

Don't keep your distance

Investigations that involve human subjects always require a close relationship between the researchers and those being studied.

The US Institute of Medicine (IOM) has, in effect, directed researchers who work with human subjects to come down out of their ivory towers. The call is made in a report, issued on 19 September, that addresses the procedures that scientists should follow when they are investigating the health of underprivileged children. But the principles that it espouses apply to a great many of those involved in publicly funded research.

The IOM study was triggered by a court case involving the Kennedy Krieger Institute, a health research centre affiliated with Johns Hopkins University in Baltimore, Maryland. The institute ran into trouble during a study on the efficacy of various techniques for removing toxic lead paint from low-cost housing. Two mothers of children in the study sued the institute, claiming that the researchers never told them that their children's high levels of lead in the blood posed dangerous health risks.

In 2001, a Maryland appeals court ruled in favour of the mothers. The court criticized ethical lapses in the study, which it compared to the infamous Tuskegee experiments on black men with syphilis. It said researchers owed a "duty of care" to the children in the study — and that the parents did not have the right to consent for children to participate, as the study was not going to benefit the children directly and could even cause some harm, through procedures such as blood testing. The institute subsequently reached an undisclosed financial settlement with the children's families.

The IOM report, *Ethical Considerations for Research on Housing-Related Health Hazards Involving Children*, makes no judgement on the Kennedy Krieger episode. But it offers some useful guidelines for managing the delicate relationship between researchers and their subjects. The Kennedy Krieger study was aimed at helping disadvantaged people, but made the subjects' families — who lived in substandard housing and had little chance of relocating — feel exploited instead. The mothers who sued the researchers claimed they were never told that their children might ingest lead, or that high blood lead levels could harm their children. The IOM points out that the

signing of a consent form wasn't sufficient to ensure that the parents understood the study.

Researchers working with disadvantaged populations should become much more involved with the communities they study, the IOM report says — the time and effort that it takes to do this will be rewarded by more convincing study outcomes. Partners in the community can help researchers, it points out, by highlighting flaws in study design, recruiting participants, and strengthening the informed-consent process.

The report also suggests that poor people in the United States are becoming more wary of participating in research, after years of involvement in studies and scant indication that the findings have any real impact on their lives.

The devastation wrought by Hurricane Katrina provided a dramatic illustration of the grounds for this mistrust. Researchers knew before the hurricane that 120,000 people in New Orleans lived in households without cars (see *Nature* 437, 174–176; 2005). Yet the city's evacuation plans made no provisions for these residents.

The IOM's recommendation that scientists work with communities to promote the translation of their findings into action applies in many different fields of study. Those who do basic research on rare diseases now realize, for example, that they must find unconventional ways to translate it into treatments, as drug companies have little interest in developing products for small numbers of patients.

Engaging communities is more difficult than simply publishing work and hoping that it will be noticed. But it is necessary, because science depends on public support, which in turn depends on the public's belief that research benefits them. If society comes to believe that researchers are operating in another world, divorced from real life, support for science can only be eroded. ■

"Poor people are wary of participating in research after years of involvement and scant indication that the findings have any real impact on their lives."

Value-free nanotech?

Efforts to gauge public attitudes to nanotechnology reveal concerns that can be readily addressed.

A profound question underlying many debates involving science and its publics is "how knowledge comes to be perceived as reliable in political settings, and how scientific claims, more specifically, pattern as authoritative".

That quote comes from *Designs on Nature* by Sheila Jasanoff (Princeton University Press, 2005; reviewed in *Nature* 437, 193–194;

2005), an overview of how the United States, Britain, Germany and the European Union have sought to deal with the issues brought up by biotechnology. Magisterial in its scope, the book takes for granted the idea, alien to many of *Nature's* readers, that science is not value-free, and that some members of the public have cultural outlooks that are simply unreceptive to accounts of what science tells us.

A corollary of this is the idea, also shared by many science-studies specialists, that attempts by scientists to communicate their discipline to the public are likely to miss the point. Only by fully engaging at the outset with the cultural preconceptions of those audiences — by being what sociologists call 'reflexive' — can science's institutions do justice to their goal of engaging with citizens. At its



worst, this agenda descends into relativism — the idea that someone's beliefs have as much weight as the so-called facts — or even Lysenkoism, in which the requirements of the state or of powerful groups take precedence over the facts. At its best, however, it can help scientists recognize how public hostility can be mobilized and consolidated despite the weight of peer-reviewed scientific evidence. This may happen because a culture of disrespect for science helps to reinforce cherished beliefs, or because experience has left individuals feeling betrayed by science or its application.

A major concern, especially in Europe, is to try to prevent such a climate enveloping nanotechnology. Experiences with genetically modified crops have led some governments to move towards being reflexive. At the same time, non-governmental organizations and other citizens' groups, more concerned about an emerging technology's potential disadvantages to their own interests, have welcomed the opportunity to tackle them as far upstream as possible.

It is in this context that two reports of citizens' participation are published this month. One, *Informed Public Perceptions of Nanotechnology and Trust in Government* by Jane Macoubrie of the Woodrow Wilson International Center for Scholars in Washington, is a study in which 177 members of the US public were briefed on nanotechnology and given a chance to explore its opportunities and their concerns. It documents weak public trust in regulatory agencies as a result of previous experiences with asbestos, dioxins, Agent Orange and nuclear power — exactly the type of cultural resistance to which Jasanoff and others would point. It also highlights a need to breed trust through better product labelling and compulsory regulation, and indicates a desire for information about the technology.

Another report, published in Britain this week, takes a less

conventional approach. Sponsored by Greenpeace, *The Guardian* newspaper and centres connected to the universities of Cambridge and Newcastle, it represents the outcome of a UK citizens' jury, in which 20 members of the public met repeatedly to hear from a variety of witnesses (see www.nanojury.org). The jury was asked about nanotechnology's benefits for the poor and disadvantaged (weak jury concerns), whether the public should determine when nanoparticles can be used in particular technologies (weak support), and whether it needs to yield more "quality leisure time" to deserve public funds (stronger support). As in the US study, the jury supported labelling, mandatory safety testing and better access to information about which nanotechnologies are being publicly funded.

These two studies reinforce the impression that the public has strong concerns about regulation and a lack of information about nanotechnology, and that nanotechnology is welcomed for its potential benefits. The results of the citizens' jury suggest that nanotechnology is not perceived as a serious threat to the values of anyone but die-hard anti-technologists. But this was a small study, and one that the jurors themselves said is provisional.

Supporters of full-blown reflexiveness should welcome a transparent citizens' jury that has probed society's assumptions about the need for a technology, and should also acknowledge these appeals for communication and regulation. Meanwhile, governments have received some direct public guidance on citizens' interests that must be protected if nanotechnology is to flourish. ■

"The results of the citizens' jury suggest that nanotechnology is not perceived as a serious threat."

Science after Katrina

The hurricane disaster on the Gulf coast will change the federal government's research priorities.

The clean-up operations in the wake of Katrina are a nightmare for all concerned, including scientists at Tulane University and other research institutions on the Gulf coast (see *Nature* 437, 177; 2005). They need all the help they can get, and other institutions must endeavour to provide it and get them back on their feet.

The ramifications of this tragedy will run deep. There are already signs that national priorities are changing: President George W. Bush, speaking in New Orleans on 16 September, broached some themes that he had previously avoided. "We have a duty to confront poverty with bold action," he said. If that pledge is to be followed through, it will involve changes in research priorities.

For better or for worse, the US federal government — particularly Congress — has a propensity to adjust the government's spending portfolio quickly in response to particular events. After the attacks of 11 September 2001, for example, the government created the Department of Homeland Security, with a large and ill-defined research programme, and diverted resources at the National Institutes of Health towards activities related to bioterrorism. Four years later, there is scant evidence that either shift has achieved much for

science or for national security. After that precedent, some would argue that the government should avoid overreacting to Katrina.

Yet Katrina has brought to the surface some critical issues that have been wantonly ignored in Washington in recent years and now deserve some attention. The most significant of these relate to poverty, as Bush has now acknowledged, and racial division. Policy-makers need good information if they are to tackle these issues, and in many instances research can provide it. The aftermath of Katrina will push poverty, at least, up the agendas of agencies such as the National Science Foundation and the National Institutes of Health that support research in the social sciences and environmental health.

At the same time, the disaster raises the profile of two very different spheres of environmental research — water management and climate change. The former is quite well understood, although a great deal of existing knowledge about rivers and wetlands, for example, is frequently ignored by policy-makers. Despite the Bush administration's scepticism about the latter, it has maintained a powerful climate-change science programme, which in time may shed valuable light on the complex relationship between global warming and extreme weather events, including hurricanes. ■

"Policy-makers need good information if they are to tackle poverty and racial division, and in many instances research can provide it."

RESEARCH HIGHLIGHTS

CANCER BIOLOGY

Persistent problem

Cancer Cell **8**, 197–209 (2005)

The recurrence of a tumour after what initially seemed successful treatment is the leading cause of death from breast cancer. A gene that plays a causal role in tumour recurrence has now been identified by Lewis Chodosh of the University of Pennsylvania School of Medicine and colleagues.

Working in mice, the team found that a gene called *Snail*, which normally helps cells change shape and migrate in embryos, was abnormally active in recurring breast tumours. Engineering tumours to express *Snail* strongly promoted their ability to recur.

Studying primary breast tumours in women, the team found that high levels of *Snail* expression were associated with an increased risk of a patient having a tumour reappear within five years. As well as being important for prognosis, *Snail* could be a useful target for cancer-fighting drugs.

GEOLOGY

Basalts flow slow

Geology **33**, 745–748 (2005)

Some 180 million years ago, the southern continent of Gondwana ruptured and the Indian Ocean began to form. Into this mighty rift flowed the massive Karoo–Ferrar flood basalts, which are preserved today in southern Africa, Antarctica, Australia and New Zealand.

Such flood basalts have been linked to mass extinction events. Earlier studies have proposed that the Karoo rocks flowed on to the Earth's surface in less than a million years — a short enough time to have constituted a massive environmental disruption.

But Fred Jourdan of the Berkeley Geochronology Center in California and colleagues think it took much longer. Using argon isotopes to date 38 rock samples taken from southern Africa, they have established that the basalts were put in place over eight million years — which might explain why there seems to be no associated mass extinction during this period.

ASTROPHYSICS

Parting stars

Astrophys. J. (in the press); preprint at

xxx.arxiv.org/abs/astro-ph/0509201 (2005)

Stars that are born together may not stay together, say Laura Gómez of the Centre for Radioastronomy and Astrophysics in Morelia, Mexico, and her co-workers.

The Trapezium is a star cluster embedded

Queen's move

Proc. R. Soc. Lond. B doi:10.1098/

rsph.2005.3234 (2005)

When a colony of social insects reproduces, queens are expected to want an equal number of sons and daughters — to maximize their genetic contribution to the next generation. But sterile female workers, being more closely related to the queen's daughters, want more females.

A mathematical model from Ido Pen of the University of Groningen, the Netherlands, and Peter Taylor of Queen's University in Ontario considers the outcome if queens and workers execute their sex-control strategies simultaneously and independently. In such a case, the ratio of males is midway between the 50% favoured by queens and the 25% favoured by workers.

But the ratio in real colonies can be nearer the queen's ideal. This is explained by a second model in which the queen acts first and the workers observe her decision. The queen places the workers in a bind by announcing her intent — once she declares that she will produce mostly males, the workers must simply favour females as much as they can. This means the queen can declare an initial position that ends up with her preferred ratio.



The queen ant must outwit her diminutive workers to produce enough sons.

in the star-forming region known as the Orion nebula (pictured). By fixing a background frame of reference for the nebula, Gómez and her team find that three radio sources in the Trapezium seem to be moving away from the spot where they were all located about 500 years ago. These objects may be young stars emerging from a multiple-star system that tore itself apart.



BIOCHEMISTRY

Predictable proteins

Science **309**, 1868–1871 (2005)

Predicting a protein's structure from its amino-acid sequence is not an easy task. Still, scientists can now calculate the structure of small proteins accurately enough to get all atoms — including amino-acid side chains — in the right places.

The predicted structures are less than 2 angstroms off the native protein structure. This is a marked improvement on previous attempts.

To get this close to the real structure, researchers led by David Baker of the University of Washington, Seattle, calculated approximate structures for the protein in which they were interested and for related proteins. The team then had enough candidate structures to start detailed atomic modelling.

So far, their method only works for proteins smaller than 100 amino acids. But with more computer power, the researchers hope to accurately predict the structures of entire protein domains.

C. KÖNIG

NASA, C. R. O'DELL & S. K. WONG (RICE UNIVERSITY)

CELL BIOLOGY

Freeing steroids

Cell **122**, 751–762 (2005)

Steroid sex hormones were once assumed to enter all cells by diffusing freely through cell membranes, modifying gene transcription only in those cells geared up to respond to them. But, puzzlingly, nearly all oestrogens and androgens circulating in the blood are bound to large proteins, and so apparently not free to diffuse.

Anders Nykjaer of the University of Aarhus, Denmark, and Thomas Willnow from the Max Delbrück Center for Molecular Medicine in Berlin and their colleagues now demonstrate the presence of a receptor called megalin on the surface of cells of reproductive tissues. This receptor causes the protein–steroid complexes to be engulfed into the cell, where the binding protein is recycled and the steroids released to do their jobs. Mice whose *megalyn* gene is knocked out have impaired sexual development.

IMAGING TECHNIQUES

A good tip

Appl. Phys. Lett. **87**, 111901 (2005)

Extremely small, short-range forces can be detected using only the thermal oscillations of a minuscule cantilever tip, say Murti Salapaka of Iowa State University, Ames, and his colleagues. The finding is useful for frequency-modulation atomic-force microscopy, which uses variations in interaction strength between a probe and a surface to non-invasively image materials at the atomic scale.

It is difficult to produce and maintain forced vibrations of the subnanometre amplitudes necessary for accurate imaging at such a scale. The thermal oscillation amplitude of the authors' probe was 0.06 nm, and they report that it could be held less than 2 nm from a sample for up to 30 minutes. This could be useful for documenting forces, such as those on cell membranes, that evolve over time.

MATERIALS SCIENCE

Shocking strength

Science **309**, 1838–1841 (2005)

Ductile metals can be made harder by making their crystalline grains smaller, because the defects that move while a metal is deforming get stuck at the grain boundaries. But this works only up to a point. Eventually, nanometre-scale grains start to slide over one another. Eduardo Bringa of Lawrence Livermore National Laboratory in California and his colleagues

may have found a way to put a stop to that.

Their computer simulations show that, by passing a shock wave through a nanocrystalline metal, they can suppress grain sliding and introduce defects into the crystal lattice that cause the grains to lock together. This should make the shocked metal harder and stronger. Preliminary experiments on nanocrystalline nickel seem to bear out the idea.

VIROLOGY

Crowning achievement

PLoS Biol. **3**, e324 (2005)

Drugs to tackle coronaviruses (pictured), which include the severe acute respiratory syndrome (SARS) virus, could be on the cards thanks to an international team led by Chinese scientists.

Coronaviruses cause a number of severe infections in humans and animals, notably in the respiratory tract. Developing effective

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

drugs and vaccines is difficult as the viruses change rapidly and species differ markedly in their protein structures. Dawei Ma of the Shanghai Institute of Organic Chemistry, Zihong Rao of Tsinghua University in Beijing, and colleagues studied the structure of a key viral enzyme, called M^{pro}, in different coronaviruses. The researchers identified a common element in this enzyme whose structure was conserved across the different viruses, then identified compounds that block it.

One compound was particularly effective at protecting cells; the team suggests that drugs developed from this compound could form a first line of defence against emerging coronavirus diseases.

JOURNAL CLUB

Sandra Knapp

Natural History Museum,
London, UK

Take a leaf out of a taxonomist's book and read about development pathways in plants.

Taxonomists in general are obsessed with the diversity of living things. The organisms we work with, be they tomatoes, dinosaurs or tiny parasitic wasps, have such a range of shape and form that they keep us endlessly fascinated.

Plants, for example, are incredible creatures. We tend to think of them in terms of flowers or fruits; those are, after all, the bits we most obviously use. But it is leaves that really drive life on Earth. And from the underwater threads of a bladderwort to the giant umbrellas of the titan arum, their diversity boggles the mind.

Finding some order in this abundance, Paolo Piazza, Sophie Jasinski and Miltos Tsiantis of the University of Oxford have reviewed the evolution of leaves in the *New Phytologist* (**167**, 693–710; 2005). They pull together the many studies on developmental pathways in model organisms such as thale cress, snapdragons, tomato and maize.

Thinking about how leaf form is regulated through wildly complex networks of transcription factors, proteins and plant hormones can be daunting. But the conceptual framework into which the work is set makes this paper accessible to those like me, whose grasp of the mechanistic details is a little bit fuzzy.

This paper gives new perspectives on my own taxonomic work on *Solanum* — with its leaves that are sometimes spiny, sometimes dissected (as in tomatoes and potatoes) and sometimes different in juveniles and adults. That does not mean I am now tempted to become a developmental biologist though. The mechanisms they review have generated much variation, which means there is a lot of taxonomy to do.

L. STANNARD/UCT/SPL

NEWS

Astronomers reject the term 'planet'

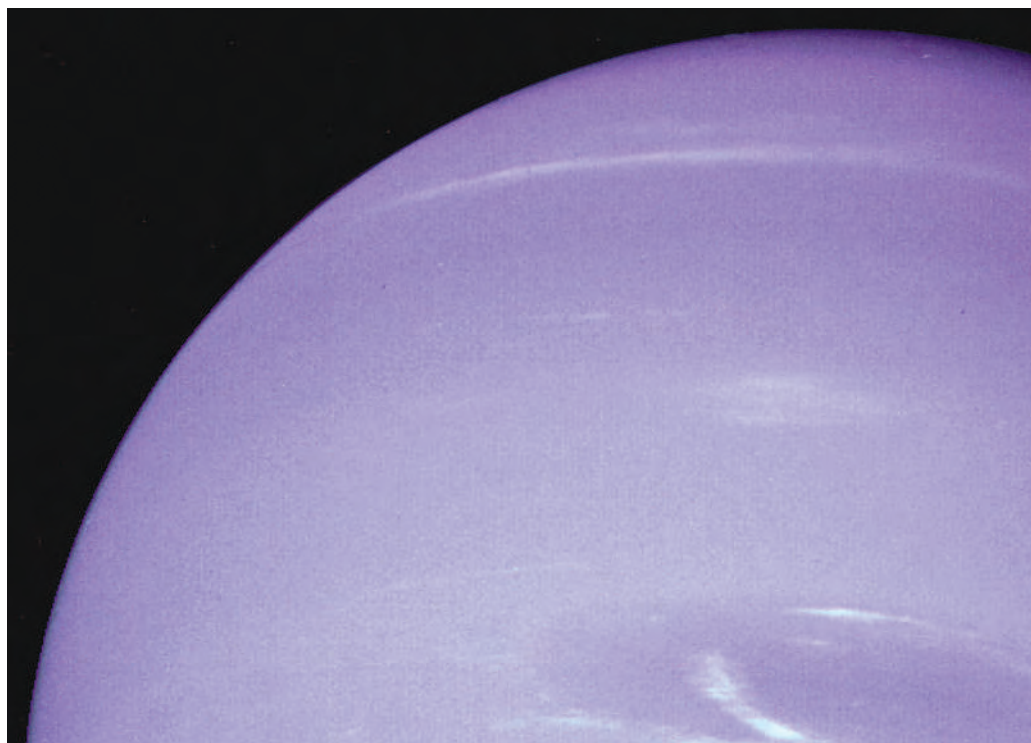
An expert panel charged with ending the debate over what is and isn't a planet has come up with a radical solution: end use of the term altogether, unless it is accompanied by a qualifier.

Debates on nomenclature are common in science, but the planet question is one of the few to have spilled into the public arena. Researchers have argued over the status of Pluto for decades, for example, with some claiming that it is not a fully fledged planet. Similar rows have raged in recent years over how to describe new additions to the Solar System. The panel could now be close to settling such matters. If it succeeds, works ranging from encyclopedias to children's books will have to be updated.

The panel's proposal, a copy of which has been seen by *Nature*, contends that the collection of objects currently dubbed planets, from rocky worlds on the outer shores of our Solar System to free-roaming objects in deep space, is too diverse to justify a single moniker. Instead, the researchers want to define different types of 'planetary object', such as terrestrial planets, including Earth, and extrasolar planets, which orbit stars other than the Sun.

"If we're going to use the word planet we should put an adjective in front of it," says Brian Marsden, a panel member and an astronomer at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts.

The 19-strong group was convened last year by the International Astronomical Union (IAU), but speeded up its work this July when a media debate broke out over the status of another addition to the Solar System. The object, known as 2003 UB313, orbits near Pluto. One of its discoverers, Mike Brown of



the California Institute of Technology in Pasadena, who has recently been courting controversy regarding another discovery (see 'Planet spotters compete'), says it should count as a tenth planet, in part because it is larger than Pluto. But other astronomers say both UB313 and Pluto are simply large members of the Kuiper belt, a jumble of rocky and icy objects that orbits some 10 billion kilometres from the Sun.

The proposal, e-mailed to group members

on 12 September, would end such arguments. UB313 and Pluto would be known as Trans-Neptunian planets, a class roughly defined as large objects that orbit the Sun beyond Neptune. Other members of the Solar System would fall into the categories of terrestrial planets or gas giants, although Iwan Williams, the group's chair and an astronomer at Queen Mary, University of London, says that his team plans only to define the Trans-Neptunian class, and will leave other definitions to the IAU.

NASA

Planet spotters compete

How to define newly discovered mini-worlds is not the only question currently dividing planet hunters. Try this: is it proper for one scientist to Google another's research? A dispute over who is the true discoverer of a Pluto-sized object at the edge of the Solar System has raised questions about the ethical use of astronomical data posted on an open website.

On 29 July, a group led by Jose Luis Ortiz of the Institute of Astrophysics in Andalusia, Spain, reported sighting an object in the Kuiper belt, about 8 billion

kilometres from Earth. The Smithsonian Astrophysical Observatory's Minor Planet Center in Cambridge, Massachusetts, which fields such claims from astronomers, temporarily designated the object 2003 EL61, based on the Ortiz group's first observation of it in March 2003.

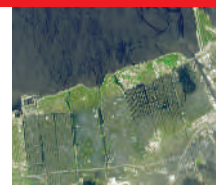
But astronomers led by Mike Brown of the California Institute of Technology, Pasadena, say they have been studying the object since December 2004, although they never announced it publicly. Brown says computer logs show that on

26 July 2005, someone accessed a public telescope database where information about his object was stored. A conference abstract written by Brown's team and posted on the Internet a week earlier contained a number that anyone with a web browser could use to access the record (*Nature* **436**, 764; 2005).

Brown has now worked out that the person who accessed the database was a member of the Spanish group, and is crying foul. He has asked the International Astronomical Union, which

oversees the Minor Planet Center, to condemn Ortiz for stealing the discovery.

Ortiz has offered no public explanation for how he used Brown's information. But he says his group first noticed 2003 EL61 in its own two-year-old data on 25 July, a day before the US group's records were accessed. In a written statement to *Nature*, Ortiz defends his actions: "If in a revision process somebody uses Google to find publicly available information on the Internet...that is perfectly legitimate." **Tony Reichhardt**



REBUILDING NEW ORLEANS' DEFENCES
The Dutch offer good advice on keeping future floods out.
www.nature.com/news

NASA

Brain imaging ready to detect terrorists, say neuroscientists

Brain-imaging techniques that reveal when a person is lying are now reliable enough to identify criminals, claim researchers.

Critics maintain that the technique will never be useful for such investigations, arguing that, as with traditional polygraph detectors, liars could learn to fool the tests. And researchers in the field have previously admitted that the approach needs more work. But neuroscientists from the University of Pennsylvania School of Medicine in Philadelphia have now told *Nature* that they believe their test is ready for real-life scenarios.

Daniel Langleben and his colleagues use functional magnetic resonance imaging (fMRI) to track people's brains when they lie and tell the truth. By analysing brain activity during both scenarios, they have developed an algorithm that can detect lies from truth with 99% accuracy.

Team member Rugen Gur points out that, unlike the polygraph, fMRI does not rely on controllable symptoms such as sweating or a fast heartbeat. Instead it monitors the central nervous system. When someone lies, their brain inhibits them from telling the truth, and this makes the frontal lobes more active. "A lie is always more complicated than the truth," says Gur. "You think a bit more and fMRI picks that up."

In the latest study (C. Davatzikos *et al.* *Neuroimage*, in the press) the team gave volunteers an envelope with two cards and \$20; subjects could keep the cash if they lied convincingly in the tests. Once they were inside the fMRI scanner, each person had to press a button to indicate whether a card flashed on the screen matched one of theirs. They were asked to be honest about having one of the

cards and to lie about having the other.

Langleben has previously warned that fMRI is a research tool, not a way to spot liars. But the latest research has changed his tune. "We can't say whether this person will one day use a bomb," he says. "But we can use fMRI to find concealed information. We can ask: is X involved in terrorist organization Y?"

The main advance is being able to distinguish lies from truthful statements in a given individual. Although previously scientists could see how the brain lit up when people lied, results were based on the averaged brain activity of a group of people and did not look at individual fibres for each person. "Now we can tell when an individual lies on a specific question," says Gur. "This is a major step forward."

Critics argue that lab experiments do not equate to real-life situations. Getting a reward for concealing a lie is not the same thing as losing your job or getting a criminal conviction for being found out, which is a far more likely consequence, says Jennifer Vendemia, an expert in lie-detection research at the University of South Carolina, Columbia. "There is nothing you can do in the lab that would mimic job loss, the death penalty, or public humiliation."

But the biggest concerns about using fMRI to detect lies, says Vendemia, are over ethical issues, such as whether individuals have the right to keep their thoughts private.

Critics and researchers agree that more funding is needed to standardize the method and iron out ethical concerns before the approach is used routinely. The team's next step is to expand its studies to include women, people of different cultures, and psychopaths. ■

Jennifer Wild

Name game: some say the word 'planet' is used too widely for it to be a useful definition.

Williams hopes to send a final version of the proposal to the IAU within two weeks, after the team has reviewed it. But whereas the broad definition of planetary objects is uncontroversial, at least one member plans to dispute the names for subtypes. "I don't believe we should classify planetary types by location," says Alan Stern of the Southwest Research Institute in Boulder, Colorado. "We should use properties of the objects as a guide."

UB313 and Pluto would be better known as "ice dwarfs", Stern suggests, because such a definition "tells us more about the objects". He points out that stars are classified by their physical properties, not their location.

If the group can reach a consensus, it will be up to the IAU's executive committee to decide whether to accept the proposal. But will the public and scientists then change the names they use for Mercury and Mars? "Old habits die hard," says Jacqueline Mitton, an author of popular astronomy books based in Cambridge, UK. She points out that some astrophysicists still describe stars as either 'early' or 'late' types, terminology that was officially abandoned around 50 years ago. "Committees can make pronouncements, but they can't always change things," she adds. "It will take a very long time."

Jim Giles



Under fire: used properly, new brain-imaging techniques might assist in investigations of suspects.

GETTY

Flu researchers slam US agency for hoarding data

Influenza researchers are complaining that the poor sharing of data by the US disease-control agency is hindering their work.

Still reeling from accusations that his administration was unprepared for the hurricane that hit New Orleans last month, President George W. Bush called last week for an international partnership on influenza that would require countries facing an outbreak to share immediately information and samples with the World Health Organization (WHO).

But investigations by *Nature* have revealed widespread concern that too few of the flu data collected by the US Centers for Disease Control and Prevention (CDC) in Atlanta are made generally available. Experts say research would speed up if the CDC's influenza branch threw open its databases of virus sequences and immunological and epidemiological data.

"Many in the influenza field are displeased with the CDC's practice of refusing to deposit sequences of most of the strains that they sequence," says Michael Deem, a physicist at Rice University in Houston, who works on predicting flu vaccine efficiency.

Policy decisions, such as which vaccine to produce ahead of each flu season, are being made without the full data being available to the scientific community, he says. "The quality of their decisions, which can affect millions of people, cannot be checked."

Deem's criticisms are echoed widely, although most scientists are reluctant to speak on the record. "This is a very delicate issue. It is important to keep a positive working relationship with the CDC, and they do lots of things well," says one evolutionary ecologist. "But getting data from them has been somewhere between extremely difficult and impossible."

Researchers say they have no idea what or even how many flu sequences the CDC processes, but it is thought to be up to thousands each year. Apart from occasional large deposits accompanying published papers, required by journals, data are "coming through an eye dropper", says one bioinformatician at the US National Institutes of Health (NIH) in Bethesda, Maryland.

Nature's analyses show that, of about 15,000 influenza A sequences in the gene database

Genbank and the influenza sequence database at the Los Alamos National Laboratory in New Mexico, fewer than a tenth were deposited by the CDC. A consortium led by the US National Institute of Allergy and Infectious Diseases (NIAID) in Bethesda has deposited more than 2,800 sequences this year alone.

"Getting data from the CDC has been somewhere between extremely difficult and impossible."

"The advancement of public health and science is generally best served when data are shared in an open, timely and appropriate manner, and the CDC Influenza Branch is committed to accomplishing this objective," says James LeDuc, director of the CDC's division of viral and rickettsial diseases. But he adds: "This must be balanced against the needs for maintaining high standards for data quality and for protecting sensitive information when the situation warrants."

LeDuc says that as well as depositing sequences alongside papers, the agency posts summaries of epidemiological data on its website each week, and shares information with the WHO. But "we do not have the capacity to comply with all requests while also meeting



Shot in the dark: a lack of data sharing could hold back the design and assessment of flu vaccines.

our other public-health responsibilities".

Many flu scientists say that the CDC should try harder. "No other US laboratory receives thousands of influenza samples and sequences from around the globe," points out one. "They

M. RIETSCHEL/AP

Industry money skews drug overviews

CHICAGO

Further evidence has emerged that money from the pharmaceutical industry is distorting the medical literature.

Meta-analyses — studies that combine the results from several trials — report more favourable results if they are sponsored by industry. The same effect had already been seen for clinical trials in several areas of medicine. But the new finding is even more worrying, say its authors, as policy-makers often give meta-analyses more weight than individual trials.

The latest result comes from a study of 71 meta-analyses of hypertension medications published between 1996 and 2002. The data in the industry-sponsored analyses were no more or less positive than those in publicly funded ones. But 93% of the meta-analyses funded by a single drug company drew positive conclusions about the medications

— compared with 79% of those from academic institutions.

Many of the industry-funded papers reached conclusions not justified by the data, says study author Veronica Yank, an expert on medical publishing at the University of Washington in Seattle. She estimates that little more than half of the industry-funded studies that reached a favourable verdict actually had the data to back it up.

"Conclusions in meta-analyses often spin the results to put them in a favourable light," says Yank, who presented her results on 17 September at the Fifth International Congress on Peer Review and Biomedical Publication in Chicago, Illinois. And, she notes, "meta-analyses often surpass clinical trials in terms of influence on policy".

"It's a marvellous study and very disturbing," says Richard Smith, chief

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

say it's in [their weekly report]," says another. "Give me a break. I want the database."

The dearth of CDC data was one reason why the NIAID last year created its consortium to sequence thousands of flu strains from

humans and birds, according to one scientist close to the project. In one of the team's first papers, published in July (E. C. Holmes *et al.* *PLoS Biology* 3, e300; 2005), researchers found that viruses swap genes with each other much more frequently than was thought.

One such swap made the virulent Fujian strain, which hit in 2003–04, and to which the annual vaccine was poorly adapted. "The minute we got our hands on some open data, it jumped out that here was something people were not aware of," says one NIH scientist. "The CDC didn't know what was going on with the Fujian thing, and by the time they realized, it was too late to use it for a vaccine."

The threat of a flu pandemic makes it "imperative that our most experienced and brilliant scientists across the globe come together as one team," says Jill Taylor, a clinical virologist at the Wadsworth Center of the New York state department of health, and a member of the NIAID consortium.

"Open data are better," agrees William Glezen, a virologist at Baylor College of Medicine in Houston. "There is a lot that we have to learn about influenza." A key issue, he says, is to match changes in the flu genome with the epidemiology of infections.

He acknowledges that CDC staff are busy with programmes such as the annual vaccine selection, and lack time and resources to share data better. "That's why other investigators need to look at the other parts," says Glezen. ■

Declan Butler



Bitter pill? Meta-analyses funded by drug companies are more likely to be positive.

executive of the London-based healthcare company UnitedHealth Europe and a former editor of the *British Medical Journal* (BMJ).

The results should alert journal staff and reviewers, admit editors. They point out that referees should pick up conclusions that go beyond the data. "This is a massive failure of peer review," says Jeremy Theobald, an editor with the publishers

John Wiley in Chichester, UK. Yank agrees: "I was embarrassed on behalf of the editors."

Yank did not say where the meta-analyses were published. Some editors say that smaller journals, which lack both the staff to scrutinize referees' reports and a large pool of submissions to choose, are more likely to accept flawed studies. Cathy DeAngelis, editor of the *Journal of the American Medical Society* (JAMA), says she is wary of industry-sponsored papers and always checks for bias, but questions whether all journals have the resources to do so.

Educating editors could tackle the problem, she suggests. Before the conference, JAMA paid for 20 editors from small journals, and some based in developing countries, to attend workshops on best practice in peer review.

Yank adds that journals can help by asking authors to place claims in the context of their data, and requiring them to own up to the limitations of the study. ■

Jim Giles



LAB LOSES TRIO OF PLAGUE MICE

Risk to public is thought to be negligible.

www.nature.com/news

ALAMY

Hurricane link to climate change is hazy

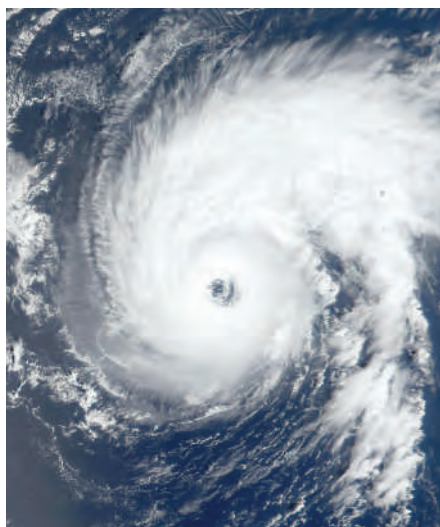
Will destructive hurricanes such as Katrina become more common in a warmer world? Two recent studies suggest that they will. But the question has split the research community (see *Nature* **435**, 1008–1009; 2005), and some say that the studies highlight how little is known about the physics of hurricane formation.

The media, and some politicians, have been quick to blame global warming for the disastrous Atlantic hurricane seasons of 2004 and 2005, including Katrina. The two new studies make no such direct link, but do suggest that the intensity — although not the frequency — of tropical cyclones is increasing sharply, perhaps as a result of global warming.

The first study, published in August (*Nature* **436**, 686–688; 2005), shows that cyclones have become more destructive over the past 30 years, with storms being both longer lived and more intense. The second, published last week (*Science* **309**, 1844–1846; 2005), concludes that today's cyclones are stronger, although less common, than 35 years ago.

"The potential for more events like Katrina is on the rise," says Greg Holland, a hurricane expert at the National Center for Atmospheric Research in Boulder, Colorado, and an author of the *Science* paper. "You can never be sure, but it seems to be consistent with global change."

But not everyone is convinced. "Hurricanes tend to go in decadal cycles," points out Chris



A brace of studies indicates that cyclones have increased in strength over the past 30 years.

Landsea, a meteorologist at the US National Hurricane Center in Miami, Florida. "Making out the tiny global-warming signal amid these large natural swings is hardly possible at this point."

And although the new studies are persuasive, they are not definitive. For instance, the *Science* paper concludes that peak wind speeds have remained constant since 1970. But this finding seems to be inconsistent with the trend towards stronger storms, says Kerry

Emanuel, a researcher at the Massachusetts Institute of Technology and author of the *Nature* paper.

Many of the arguments arise because there are insufficient data. "There seems to be a monumental lack of understanding of the role hurricanes play in the climate system," says Peter Webster, an atmospheric researcher at the Georgia Institute of Technology in Atlanta and an author of the *Science* paper.

One uncertainty, he says, is how changes in sea surface temperature affect the formation of tropical cyclones. Hurricanes in the Atlantic Ocean, for example, seem to respond to such changes quite differently from typhoons in the Pacific. But just 12% of cyclones form in the Atlantic, the only region where reconnaissance aircraft fly regularly to make precise measurements of hurricane winds.

Research is now under way to address the issue. As a follow-up to the *Science* paper, the authors have started to investigate how oceans interact with the atmosphere in different parts of the world. The goal, says Judy Curry of the Georgia Institute of Technology, is to determine how and why natural fluctuations favour different hurricane patterns in different ocean basins.

Understanding these complex mechanisms, she says, should help researchers better quantify how much a storm trend is due to global warming — or not.

Quirin Schiermeier

J. DESCLITRES, MODIS LAND RAPID RESPONSE TEAM/NASA/GSFC

Lack of lab notes casts doubt on RNA researcher's results

TOKYO

A respected Japanese scientist who failed to produce laboratory notebooks confirming his published results now faces a furore over the credibility of his findings.

On 13 September, the University of Tokyo's School of Engineering held a press conference to say that Kazunari Taira, a professor at the school who specializes in RNA research, had not provided raw data to verify his team's results. The RNA Society of Japan has also questioned some of Taira's methods.

Last year, the RNA Society of Japan began receiving letters from scientists in Japan and elsewhere saying that they could not repeat Taira's results. In April the society asked the University of Tokyo to examine 12 of Taira's papers published between 1998 and 2004.

The university set up a committee of

internal and external experts to examine four papers out of the 12, including two published in *Nature* (H. Kawasaki and K. Taira *Nature* **423**, 838–842; 2003 and *Nature* **431**, 211–217; 2004) — the first paper had already been retracted (*Nature* **426**, 100; 2003) and the second corrected (*Nature* **431**, 878; 2004). The panel asked Taira to submit samples and notebooks relating to the experiments, but the researcher in his lab who ran the experiments did not have them. The researcher had stored data in a computer, some of which had since been scrapped.

The university press release said that "the investigation committee so far could not confirm the credibility of research results because it could not confirm the existence of clear data to support those results". It has asked Taira to do the experiments again, and

will produce a final report by March 2006.

Taira says that not taking notes was "not common sense" and was regrettable. All the other researchers working with him keep notes, he says.

But Taira says that the oversight does not mean his methods are wrong, and says other groups have used his technique to publish findings. He also says that other researchers' notebooks back up some of the experiments. He now requires his researchers and students to get their notes signed by a third party.

Kimihiko Hirao, head of the School of Engineering, says that, unlike many countries, Japan doesn't have independent bodies to monitor scientific wrongdoing. "It's time to consider establishing a third-party regulatory system," he says. **Ichiko Fuyuno**

The research vessel *Nancy Foster* has been sampling the Gulf of Mexico to gauge flood pollution.



NOAA

After Katrina: tracking the toxic flood

BATON ROUGE

Three weeks after Hurricane Katrina ravaged the coasts of Mississippi and Louisiana, marine researchers are starting to assess the safety of fish and shellfish exposed to toxic flood waters in the Gulf of Mexico.

The flood waters are teeming with *Escherichia coli* bacteria and a wide range of chemicals (see *Nature* 437, 301; 2005). And engineers are pumping the toxic mix out of the city towards the Gulf coast. With its shrimp, oyster, crab and flat-fish stocks valued at around \$3.1 billion, the coast is one of the richest fishing grounds in the United States.

Shailer Cummings of the National Oceanic and Atmospheric Administration led a food-safety team into the Gulf last week for a three-day expedition on board the research vessel *Nancy Foster*. Until then the boat was being used to check the safety of the region's major ports, doing soundings to check for obstructions under the water, among other tasks.

Cummings had just two days to organize the expedition — a huge challenge under normal circumstances and more so as personnel were in such short supply. “When you try to get a research team from a distressed area they are hungry, stressed and looking for their family like everyone else,” he says. So he recruited a team from the Northwest Fisheries Science Center in Seattle, Washington.

His team of some 15 scientists collected shrimp, oysters and Atlantic croakers — a

common ground fish — and sent them to a Seattle laboratory to be tested for bacterial contamination and pollutants. Aware of the risk of a second public-health disaster in the wake of the hurricane, the researchers worked around the clock. “Nobody sleeps,” Cummings told *Nature* after three tough days aboard the *Nancy Foster*. “We’re doing this to make sure the food supply is safe.”

On their tour through the Gulf, the crew took samples from the muddy plume that has spread from the Louisiana coast over hundreds of square kilometres. Katrina muddied the water by washing sediments from the Mississippi into the sea, explains Cummings. And the fresh water is sitting on top of the salty water, spreading “like tea on a tabletop”.

So far, the team has seen no evidence of the algal bloom that might result from the freshwater influx, and the circulation pattern of the water seems normal. But “we don’t know what’s in the water”, says Cummings.

As well as sampling water and sediment for pathogens and chemicals, the researchers dissected, prepared and labelled fish samples for storage until they could be tested back in Seattle. It was close, feverish and smelly work. The liver and bile will be tested for fat-based contaminants such as polyaromatic hydrocarbons. The gut and gill will be tested for pathogens, and the muscle will be tested

for mercury and other contaminants.

“It’s an unprecedented situation,” says Tracy Collier, head of the ship’s wet lab and director of Environmental Conservation at the Northwest Fisheries Science Center. “We’re trying to sample as broadly as we can,” he explains, to detect anything that might harm people.

The results will not be ready until the end of September. But it is unlikely anyone will get ill in the meantime, says Bo Boehringer of the Louisiana Department of Wildlife and Fisheries, because no fishing is likely to take place for many weeks. Katrina has brought the

fishing community in the region to a standstill. Boats and piers have been destroyed, ice houses smashed and fish-processing centres damaged. US authorities estimate that about 4,800 fishermen

in the area are now out of work.

Although serious attention is being paid to food safety, there is also concern that Katrina has damaged fish and shellfish stocks. Flood water has covered oyster beds along a stretch of inland bays with up to a metre of mud, and the oysters have suffocated.

Some wildlife seems to be bouncing back, however. Biologist Melody Baron was on board the *Nancy Foster* to watch for marine mammals in distress. On her third 12-hour shift she said she had seen turtles and dolphins behaving normally.

Adrianne Appel

“We’re making sure the food supply is safe.”

Race claims spark fury over Croatia's school curriculum

Croatian scientists are angry. They believe their science minister is getting schoolteachers to promote the view that Croats are only distantly related to other Slavic populations such as the Serbs.

The claim that there are racial differences between Croats and Slavs is not accepted by geneticists and is potentially incendiary in the Balkan region, recently torn apart by civil war.

Dragan Primorac, minister of science and education, who trained as a medical scientist, denies that he wants the idea taught in schools. "We need much more scientific evidence before we draw conclusions," he told *Nature*.

But Primorac has regularly upset scientists over the past two years by claiming in newspaper and television interviews that Croats are descended from ancient civilizations that migrated from mid-Asia. The reports quote him as saying that Croats are more similar to Finns than other Slavs.

Last week, discontent boiled over when one of Primorac's advisers, Vladimir Paar, was interviewed by the newspaper *Jutarnji List*. Paar is a physicist from the University of Zagreb and coordinates the national school curriculum for Primorac. He said that the pilot phase of a new curriculum has just started in 5% of Croatian schools and includes history classes on the use of scientific techniques, such as genetics, to analyse the distant past of human populations.

In the article, Paar referred to Primorac as "the most competent person in the world" and said: "It is good that kids see that scientific investigations can contribute to human

history. Very soon it will be confirmed that Croats are among the oldest nations in Europe and that Hungarians carry more Slavic markers than Croats." However, he told *Nature*, specific examples like this would not be included in textbooks.

Primorac's arguments are based on data he provided for a *Science* paper (O. Semino *et al.* *Science* 290, 1155–1159; 2000) that looked at a series of genetic markers on the male, Y chromosome of human populations across Europe. The aim of the study was to determine the genetic legacy of palaeolithic *Homo sapiens*. Primorac has since been reported as saying that the markers in his samples show that Croats are more closely related to Germans and Lapps than to Slavs. These conclusions were not drawn in the *Science* paper itself.

Geneticists argue against his conclusions on several grounds. For example, only one out of the dozen or so genetic markers analysed revealed that Croats are more similar to northern populations than to other Slavs. "Croatia has a profile of genetic markers similar to the rest of the Balkans," says Ornella Semino, a geneticist at the University of Pavia, Italy, who led the *Science* study.

Scientists inside Croatia are cautious about engaging in public criticism. But Miroslav Radman, a geneticist born in Croatia and now working at the Necker Institute in Paris, is blunt. "The selection of data to support Primorac's conclusions is appalling," he says. "And to extrapolate so insensitively from sparse evidence is irresponsible."

Alison Abbott

Child care: the teaching of genetics is sensitive in Croatia, given the region's history of racial tension.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

CLEO PHOTO/ALAMY

ON THE RECORD

"Delaying having children defies nature and risks heartbreak."

The *British Medical Journal* warns women who want children not to defer pregnancy.

"You have more security at a McDonald's than at some of these facilities."

Microbiologist Richard Ebricht criticizes security measures at a New Jersey lab that has lost three plague-infected mice.

"Some recipes call for 30 frogs for a single dish."

Biologist Esteban Lavilla describes how local customs are causing frog populations to dwindle in parts of South America.

Sources: *Br. Med. J.*; Newark Star-Ledger; BBC

SCORECARD



Big brother

Dutch authorities plan to track every newborn from cradle to grave — collecting health, family, school and police data in a single electronic source.



Occupational hazards

Office drones and grad students aren't alone: Brazilian farmers who hand-milk their cows risk carpal tunnel syndrome too.



Be prepared

US security secretary Michael Chertoff is criticized for attending a bird flu meeting while his agency fumbled to cope with Hurricane Katrina. At least he's ready for one impending disaster.

NUMBER CRUNCH

96% of Americans say it is important to invest in medical research to provide a solid scientific base for health care.

73% of them do not know that the National Institutes of Health is the agency mainly responsible for funding medical research.

18% say they know a scientist personally.

Source: *J. Am. Med. Assoc.*

W. WALLAUER



Saved? An international declaration aims to secure the future of wild great apes by 2015.

African nations rally to safeguard great apes

Hoping to save our closest kin from extinction, delegates at the first Intergovernmental Meeting on Great Apes signed a declaration on 9 September pledging to secure the future of all species and subspecies of great apes in the wild by 2015.

The signatories of the "Kinshasa Declaration", who gathered in Kinshasa, the Democratic Republic of the Congo, include the governments of many of the countries where the apes live, as well as donor governments, conservation organizations, the United Nations Environment Programme (UNEP) and the United Nations Educational, Scientific and Cultural Organization (UNESCO).

UNEP estimates that less than 10% of the apes' original forest home in Africa will be left by 2030. The declaration explicitly calls for poverty reduction and the promotion of sustainable livelihoods as a means of saving the apes, which are especially threatened by poaching and habitat clearing.

Mars Express gets boost despite instrument failure

There was both good and bad news last week for Europe's Mars Express spacecraft, which has been orbiting the Red Planet for nearly two years.

On 15 September, the European Space Agency's science programme committee decided to keep operating the spacecraft until at least November 2007. That's a two-year extension on its mission to study the martian surface and atmosphere.

But Mars Express also suffered a serious blow: the Planetary Fourier Spectrometer (PFS), one of seven instruments on board the orbiter, has been shut down after behaving erratically for months. The PFS measures carbon dioxide and trace elements in the martian atmosphere, and its detection

of elevated methane levels led some scientists to speculate about biological activity on Mars.

Without a working PFS, the source of the methane will be hard to pin down. A team of engineers from ESA, the Italian Space Agency and industry is looking into the problem.

NASA sketches out plans for return to the Moon

NASA has announced specifics of how it plans to return astronauts to the Moon, a vision laid out by President George W. Bush that is expected to cost some \$100 billion.

Under the plan, a lunar lander and booster would be launched into orbit atop a massive cargo rocket by 2018. A four-person crew would launch separately in a yet-to-be-designed crew exploration vehicle. Both vehicles would be based on existing shuttle technology.

The crew would rendezvous with the booster in Earth orbit and then travel to the Moon. In announcing the plan on 19 September, NASA administrator Michael Griffin said that the agency would pay for the vehicles using its existing budget and infrastructure.

Researchers say security law restricts freedoms

US scientists are challenging Department of Defense (DOD) rules that could restrict the flow of information in federally funded projects involving foreign nationals.

The DOD has long limited the access of foreign nationals to certain information and technologies on the grounds of national security. But only recently did the department learn that some DOD contract holders, including several universities, were unaware of these laws.

In July, the DOD proposed guidelines to reinforce the existing laws, but academics have spoken out against them. Critics note that the rules would require separate work

areas and badges for certain projects involving scientists from other countries. The critics argue that such rules would limit academic freedom.

"Enforcing the controls would be nearly impossible," wrote Vartkess Apkarian, chair of the chemistry department at the University of California, Irvine, in a comment sent to the DOD. "The only choice I would have as a department chair is to reject any foreign student subject to" the laws.

At a workshop at the National Academies last week, DOD officials said the rules would probably change in response to comments received by the deadline, which has been extended to 12 October. Comments can be submitted by visiting www.acq.osd.mil/dpap/dars/publiccomm/index.htm

Creativity wins the day in US prize awards

A pharmacist, a lobster fisherman and a violin-maker are among this year's 25 recipients of MacArthur fellowships, a set of US\$500,000 grants awarded to promising creative individuals.

Pharmacist Michael Cohen, president of the Institute for Safe Medication Practices in Huntingdon Valley, Pennsylvania, has championed changes in drug naming to help avoid patients being accidentally given the wrong prescription. Lobster fisherman and biochemist Ted Ames, based in Stonington, Maine, aims to combine scientific analysis with his decades of hands-on experience to develop new strategies for managing fisheries on the US east coast.

Also featured on this year's list are Joseph Curtin, a violin-maker based in Ann Arbor, Michigan, who is seeking to use twenty-first-century materials and techniques to build better instruments, and Claire Gmachl, a laser technologist at Princeton University whose work could lead to new techniques for environmental monitoring and clinical diagnosis.

Film project tracks butterflies across North America

Stay tuned for the latest animal film stars: monarch butterflies. This ultralight airplane, dubbed Papalotzin, is following and filming the migration of monarchs across North America this year.

Monarchs are approaching the peak of their annual migration, which takes them from Canada and the northern United States to California and Mexico for the winter. Conservationists hope that the film, being shot from the two-person ultralight, will raise awareness of the need to save ecosystems along the butterflies' flight path.

► www.papalotzin.com



WWF

LEFT BEHIND

Two researchers survived the worst of Hurricane Katrina, caring for sick patients in a flooded hospital. **Erika Check** hears of their harrowing experience.

As Ruth Berggren slept in a darkened hospital room, the hot, muggy air settled damply on her skin. Nothing stirred; the air-conditioning and lights had shut down days before, when Hurricane Katrina knocked out Charity Hospital's main power. Berggren and her husband Tyler Curiel, leading scientists in New Orleans' academic community, had been waiting for rescue for days. They were hoping to save the patients of Charity's Ward 9 West.

Suddenly, Berggren heard someone shouting. She leapt from her bed and ran into the pitch-black hallway. She saw a sweaty young Marine, twitching nervously as he held his gun. "What are you doing here?" he barked, shining a flashlight in her face. "I was told this floor had been evacuated!"

Hearing the commotion, Curiel bounded into the hallway. The soldier, startled, raised his gun. "I'm the physician in charge of this ward, and we have not been evacuated!" a terrified Berggren shouted. The Marine lowered his gun and continued on his rounds.

Later, Berggren recalled the episode as another frustrating reminder of her patients' situation. In the midst of a lawless, desperate city, poor people with AIDS had been abandoned yet again.

Berggren and Curiel, both faculty members at Tulane University, landed at the epicentre of the Katrina disaster more or less by accident. Berggren, an infectious disease specialist, teaches at Charity Hospital one month a year. She happened to be on call on Sunday 28 August, the day Katrina's winds began to pound the US Gulf Coast. Curiel, a cancer immunologist and oncologist, stayed behind with her.

The first days were relatively calm. The storm blew in on Sunday morning and out by Monday evening. At Tulane's medical centre, across the street from Charity, Curiel cared for patients and worried about what might happen if the power failed. He and a graduate

"A man in a white shirt was firing bullets at the rescue workers from a parking deck."



J. CHAMBERS



Flight to safety: researchers Ruth Berggren and Tyler Curiel endured days in the flooded medical complexes of New Orleans (left).

student rushed through his lab, transferring valuable cell lines from electric freezers into tanks cooled by liquid nitrogen.

By Tuesday morning, New Orleans was drowning, swamped by waters that poured through a levee breach. Administrators at Tulane, a private hospital, hired helicopters and evacuated their patients. Then Curiel borrowed a canoe and paddled across the street to his wife.

The situation at Charity was disintegrating. The hospital had no money to call in a private evacuation for its 250 patients. Bottled water was plentiful, but food was rationed to a few cups of ravioli and a handful of canned green beans each day. When the hospital ran out of diesel to power its generators, the staff helped patients on respirators breathe with hand pumps.

Despite the worsening conditions, the Ward 9 West shift manager, Mitch Handrich, vowed that no staff members would leave until every patient was evacuated. The staff tried to bolster patients' spirits; on Wednesday, they staged a talent show for the ward. That night, the National Guard finally arrived and evacu-

ated five of Berggren's patients. But on Thursday morning, gunshots erupted as Curiel was helping load patients into trucks on the emergency room ramp. Curiel says he saw a man in a white shirt firing bullets at the rescue workers from a parking deck. "Sniper!" the guardsmen shouted.

Toxic smoke

Staff members scrambled to get the patients to safety. But the damage was done; the National Guard called off the operation and abandoned the building. Only 50 patients had been evacuated, leaving 200 stranded — including 13 in Berggren's ward. That night, Berggren and Curiel had their encounter with the Marine. When they woke up Friday morning, a chemical plant had exploded nearby, spreading a pall of toxic smoke for miles around. For the first time, they began to think they might not make it out of Charity alive.

But television coverage of the shootings alerted the nation to the hospital's plight. On Friday, the National Guard relaunched its evacuation effort. All of Berggren's patients were out of the hospital by Friday afternoon. The next day, Curiel, Berggren and her ward's staff evacuated to Texas.

One week later, Curiel returned, escorted by armed guards, to learn the fate of his research materials. He thought his life's work had been lost. But when he opened the first nitrogen tank, a tell-tale puff of cool vapour wafted up into his face. He knew then that his samples were safe.

Erika Check is a Washington correspondent for *Nature*.

Inside information

Earth's climate depends strongly on clouds. But what really goes on within these layered structures? **Heike Langenberg** reports on two satellites that aim to find out.

There's more to clouds than meets the eye. Look up on an overcast day, and you'll be presented with a sea of grey. But this blanket blotting out the Sun is merely the base of an intricate layered structure that rises vertically through the atmosphere.

The composition of these layers helps to dictate what effect the clouds have on our climate. The results can be dramatic — but little is known about how these effects come about. Two satellites, scheduled for launch next month, are set to change all that.

Named CloudSat and CALIPSO, these satellites will use radar and an equivalent technology based on light waves, known as lidar. They will cut through the layered structure of clouds to see how water droplets and airborne particles, or aerosols, are distributed around the globe.

"This is a truly exciting time," says Graeme Stephens, a climatologist at Colorado State University and head of the CloudSat science team. "We're entering a new era of Earth observations with these missions."

Clouds are one of the last great unknowns when it comes to understanding Earth's climate. They can both absorb and reflect the Sun's radiation before it reaches the planet's surface. They can also capture outgoing radiation from Earth. The scope of this effect depends in part on the distinct layering of the clouds and the variations in colour, density and altitude.

The small water droplets that constitute

clouds, and the even tinier particles that make up aerosols, act on scales much smaller than the 100-kilometre grids typically used in climate model calculations. Even worse, the distributions of clouds and aerosols can change rapidly over time, which limits the usefulness of observations taken during patchy measurement campaigns.

The Intergovernmental Panel on Climate Change recognized such uncertainties in its most recent report, citing cloud and aerosol effects as among the least understood of the factors that affect Earth's climate.

Cover story

But clouds are too important to climate for them to be ignored. A change of only about 1% in global cloudiness can either mask or double the effect that a decade's worth of greenhouse-gas emissions have on the amount of heat lost to space from Earth, says Bruce Wielicki, a climate researcher at NASA's Langley Research Center in Hampton, Virginia.

And human activity can bring about changes in global cloudiness. Global warming caused by greenhouse-gas emissions can change the planet's water cycle, for example, and tiny aerosol particles emitted in pollution can change cloud properties and precipitation.

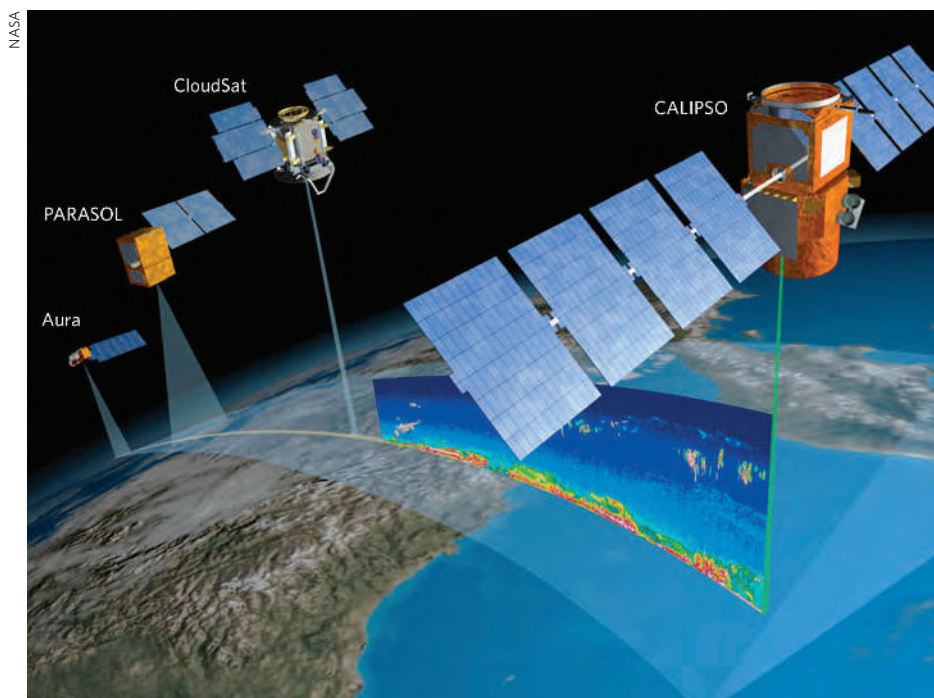
Researchers continue to debate whether climate change has already brought about long-term changes to incoming solar radiation by altering cloudiness and so leading to either a

dimming or a brightening on a global scale^{1–3}. Such uncertainties are rooted partly in a lack of continuous, accurate cloud measurements from around the planet.

Until now, most observations have measured only total cloud thickness. The new missions will go much further. One of CALIPSO's tasks, for instance, will be to observe the monthly global mean cloudiness down to variations of just 0.1%, says David Winker, who heads the satellite's science team at the Langley Research Center.

Modelling cloud and aerosol dynamics presents problems for climate prediction, as the uncertainties are difficult to constrain⁴. Until now, getting data on the vertical distribution of cloud layers has been restricted to occasional reconnaissance flights. As a result, climate researchers cannot accurately estimate how aerosols will affect the radiative properties of clouds, a process known as the first indirect aerosol effect. Current satellites cannot tell whether or not aerosols and clouds occur at the same altitude, and so cannot tell whether the two are truly linked.

With its 3-millimetre wavelength radar, CloudSat is designed to detect relatively large water droplets in both thick and thin clouds. CALIPSO — a tortured acronym for Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation — will use much shorter wavelengths to distinguish the fine details of thinner clouds and aerosols. Combining this



When they launch, CloudSat and CALIPSO will join a group of Earth-observing satellites called the A-train.

information should provide a new benchmark for existing climate models, says Ulrike Lohmann, a climatologist at the Swiss Federal Institute of Technology in Zurich.

The two satellites will fly in formation about 700 kilometres above Earth's surface, and will measure the same spot in the sky within 15 seconds of each other. These near-instantaneous observations will allow them to identify when clouds and aerosols are truly in the same layer, as opposed to located in separate layers that just happen to be vertically aligned.

Making a splash

CloudSat will also observe rain, so that researchers will similarly be able to tell when precipitation occurs in the same air mass as aerosols and cloud droplets. That, in turn, will help them to investigate the second indirect aerosol effect — how aerosols affect precipitation and cloud lifetime on a global scale⁵. "CloudSat will be like a medical scan revealing the inner workings of clouds," says Stephens.

Unlike many satellites, CloudSat and CALIPSO carry 'active' instruments, which send out their own electromagnetic signals — radio waves for CloudSat's radar and light waves for CALIPSO's lidar. These instruments record the time it takes for the signal to bounce off an obstacle, such as a cloud, and return to the satellite, thus revealing the cloud's altitude.

CloudSat's radar will be able to cut through all but the densest clouds. The shorter-wavelength lidar on CALIPSO cannot cut through thick or frontal-storm clouds, but should be able to profile about 60% of all clouds, says Winker.

The idea of using both radar and lidar to study clouds and aerosols from space has been around since at least the early 1990s, Stephens

notes. But putting both instruments on one satellite proved too costly. By splitting the project in two, NASA spent \$200 million on CloudSat and \$175 million on CALIPSO. France's space agency helped to build the latter, and Canada contributed a key element to CloudSat's radar.

Both satellites will be launched on a single rocket no earlier than 26 October from the Vandenberg Air Force Base in California. They will join three Earth observation satellites already in place: Aqua, which investigates

"The statistical information from satellite data is invaluable for the validation of climate models."

— Ulrike Lohmann

Earth's water cycle; Aura, which focuses on air quality and stratospheric ozone; and PARASOL, which measures the direction and polarization of light in the atmosphere.

Together, the satellite formation is called the 'afternoon constellation', or the 'A-train', because it crosses the equator at about 1:30 p.m. local time. One final satellite, the Orbiting Carbon Observatory, is slated to join the A-train in 2008. With only about 8 minutes between the passage of Aqua, the first satellite in the A-train, and Aura, the formation's tail end, the instruments see largely overlapping portions of the sky.

"The A-train as a whole will sample practically all the clouds above Earth," says Stephens. Aqua's observations of cloud optical properties, in particular, should be invaluable in interpreting CloudSat's information on cloud layering.

CloudSat and CALIPSO will also complement each other, as aerosols are intricately linked to the formation of clouds. In aerosol-rich air, for instance, clouds tend to be composed of a large number of small water droplets — a phenomenon that can affect the brightness of the clouds as well as their capacity to produce rain.

Not all aerosols have the same effect. Most aerosols simply help cloud droplets to condense, but soot particles can also absorb some of the Sun's radiation and so warm the atmosphere, potentially thinning the cloud layers. CALIPSO will be able to distinguish the size and shape of aerosol particles. As those from human sources are typically ten times smaller than particles from natural sources, this should allow researchers to track specific sources of pollution.

Arctic role

Another area in which the satellites might contribute is charting Arctic clouds. To 'passive' instruments, which don't send out their own radar or lidar signals, these bright, cold clouds are hard to distinguish from the icy surface they float above. The active instruments on CloudSat and CALIPSO should be able to get around this problem by resolving these clouds into three-dimensions — making them stand out against the white background much like objects in a 3-D film do when the viewer dons red-green glasses.

CloudSat and CALIPSO's new observations should help modellers to see whether their simulated clouds are not only over the right region but also at the correct height. Repeating measurement slices through similar atmospheric conditions should uncover ways to formulate the model equations and, eventually, create more realistic models. "The statistical information from satellite data is invaluable for the validation of climate models," explains Lohmann.

Yet for all their strengths, the two satellites cannot solve all the research problems of clouds and climate. Ideally, scientists would be able to trace air masses as they move, in order to study the evolution of clouds in the presence or absence of aerosols. CloudSat and CALIPSO will capture only a frozen picture in time, rather than the entire dynamic movie.

But for now, climatologists are excited about the prospect of getting so much fresh information and finally plugging some holes in their understanding of Earth's climate. For once it seems it's not a bad thing to have your head in the clouds.

Heike Langenberg is a physical sciences editor at Nature.

1. Wielicki, B. A. *et al. Science* **308**, 825 (2005).
2. Wild, M. *et al. Science* **308**, 847–850 (2005).
3. Pinker, R. T., Zhang, B. & Dutton, E. G. *Science* **308**, 850–854 (2005).
4. Andreae, M. O., Jones, C. D. & Cox, P. M. *Nature* **435**, 1187–1190 (2005).
5. Ackerman, A. S., Kirkpatrick, M. P., Stevens, D. E. & Toon, O. B. *Nature* **432**, 1014–1017 (2004).

Back to school

This month, as most researchers gear up to teach, two scientists are heading into the classroom to learn.

Geoff Brumfiel asks why a physicist would want to enrol in biology lessons.

This spring, David Hartley found himself leading a double life. By day, he was teaching a class at the University of Maryland School of Medicine in Baltimore. By night, he was a student at George Washington University in Washington DC.

Hartley, a surface physicist by training, is one of a small but growing number of professional physical scientists who are going back to university to begin new careers in biology. Under a unique grant programme from the National Institutes of Health (NIH), about 20 mid-career scientists each year are going back into the classroom. The grants, known as K25s, allow researchers to spend up to five years learning the discipline of their choice.

"Physics has always been the centre of my world, ever since I can remember," Hartley says. But by next spring, he will have completed a master's degree in public health and will begin an active research programme that will combine his knowledge of maths with his new-found love for epidemiology.

Even second time round, university can be daunting. At 37, Hartley has found juggling the roles of student and researcher particularly testing. "It has been very taxing in the past year, trying to do research, teach, do coursework and maintain a life with my family," he says.

His new student life has required a shift in thinking as much as a shift in routine. Trying to grasp the breadth of a field such as epidemiology is tricky, Hartley concedes. "You have sociologists, statisticians, epidemiologists and microbiologists all working in public health — and they don't necessarily speak the same language," he says. "The most challenging thing for me is getting my arms around all of this."

Since the K25 programme began in 2000, the NIH has given out more than 120 grants to chemists, engineers, physicists and mathematicians, all of whom are looking to transfer their skills to biological research. From the agency's perspective the potential benefit is

huge. Key developments in the biological sciences, such as protein crystallography, electron microscopy and magnetic resonance imaging, were all pioneered by physicists. And increasingly, biomedical science is being advanced by broad, interdisciplinary teams of diverse scientists, according to Norka Ruiz Bravo, the NIH's deputy director of extramural research.

Hartley, who had spent years in private industry doing defence-related research, saw the K25 grant as a route back into academia.

"My first reaction was, oh God, these guys are so immature, but then I stopped and thought, no, I was just like that." — David Beebe

But for a scientist with the security of a faculty position, what is the motivation to switch fields in mid-career? For David Beebe, another K25 participant, it was a combination of professional interest and personal experience.

After training as an electrical engineer, Beebe had become, at 41, a successful researcher in microfluidics, the flow of fluids through tiny channels. His 25-person lab at the University of Wisconsin at Madison was churning out about a dozen papers a year, and he was considered a leader in the field.

Switching subjects

Among other projects, Beebe began to experiment with ways that microfluidic channels might be used to automate certain aspects of the *in vitro* fertilization process. "We started doing some experiments and saw that, lo and behold, the embryos were developing almost as quickly as they were *in vivo*, but we didn't know why," he says. "I realized I needed to know more biology."

Around the same time, Beebe says, he found himself becoming more personally drawn to biology. A friend had developed breast cancer, and that got him thinking about how his research might be able to contribute to advances in medicine. "I really wanted to make a contribution to humankind, as silly as that sounds," he says.

So Beebe decided to go back to school and

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

S. GREENHILL & R. GREENHILL/ALAMY

study cancer biology. Originally he wanted to apply for a full PhD programme, but he was dissuaded by colleagues. "Most of the people here said, 'You're crazy, why do you want to get another PhD?'" he explains. Ultimately, he chose to focus on classes that interested him rather than trying to earn another degree.

Once back in school, Beebe faced a culture shock. In his first year, he found himself in an undergraduate-level organic chemistry course, and was horrified to discover that many of his 300 classmates were more interested in discussing parties, sports, boys, and anything other than the organic molecules they were there to study. "My first reaction was, oh God, these guys are so immature," he says, "but then I stopped and thought, no, I was just like that." Beebe says that the experience has reminded him "where his students' heads are at" when they come to his classes.

These experiences have also helped him to relate better to the students in his lab, according to Jaisree Moorthy, a postdoc and former graduate student of Beebe's. This is despite worries that the K25 programme would be disastrous for the lab. "I had 20 graduate students, and only found out I got the grant a few months ahead," Beebe recalls. "So ramping down has really been the biggest struggle." But the programme has actually made Beebe more available to his students: "Since Dr Beebe has gone back to school, he is more accessible as he does not travel as much," says Moorthy.

Despite their overwhelming schedules, colleagues say that both researchers are already making significant contributions to their adopted fields. "David Hartley really asks some interesting research questions," says Eli Perencevich, an epidemiologist at the University of Maryland. Last year, colleagues asked Hartley, Perencevich and others to study an outbreak of methicillin-resistant *Staphylococcus aureus* that had occurred in the university hospital's ward for prisoners. "David started thinking about what a person who wanted to control an infection would need to know," Perencevich says. "And he was able to take a very complex model and break it down into three or four parameters that infectious-health professionals could use to control it."

Beebe's studies could have payoffs in the field of stem-cell research, according to his mentor Caroline Alexander, a geneticist at the University of Wisconsin. Alexander and Beebe have been collaborating on a mouse model of how breast cancer tumours develop. They believe that the origin of the tumours may be linked to stem cells in the mouse's breast. "It's become very interesting from a therapeutic point of view to understand the biology of these stem cells," she says. Researchers have long wanted to study how proteins secreted by these cells regulate their development. But this is almost impossible to do in a normal Petri dish because it contains millions of cells all signalling to each other at once, Alexander says.



Lab life: geneticist Caroline Alexander helps David Beebe, a microfluidics researcher, to sharpen up his skills in cancer biology.

By contrast, the narrow channels in Beebe's microfluidic device hold far fewer cells, and their environment is easier to control. By switching the fluid flow through the channel on and off, the researchers can modify the signalling activity. "This is a great environment to test out how cells are regulating one another," Alexander says.

The road ahead

When Hartley finishes his courses next spring, his research will kick into high gear. Rather than a thesis, the second half of the K25 programme requires him to produce publishable research findings. Hartley plans to develop a mathematical model linking cholera outbreaks to environmental, social and climatic factors. Beebe says that after his courses end next autumn, he hopes to begin asking his own research questions, rather than

"For us to make progress, we've got to recognize the importance of people from alien disciplines."

— Glen Morris

following the cues from his collaborators. "Next fall, I'll take complete ownership of my research," he says. "To me, that's going to be the big transition."

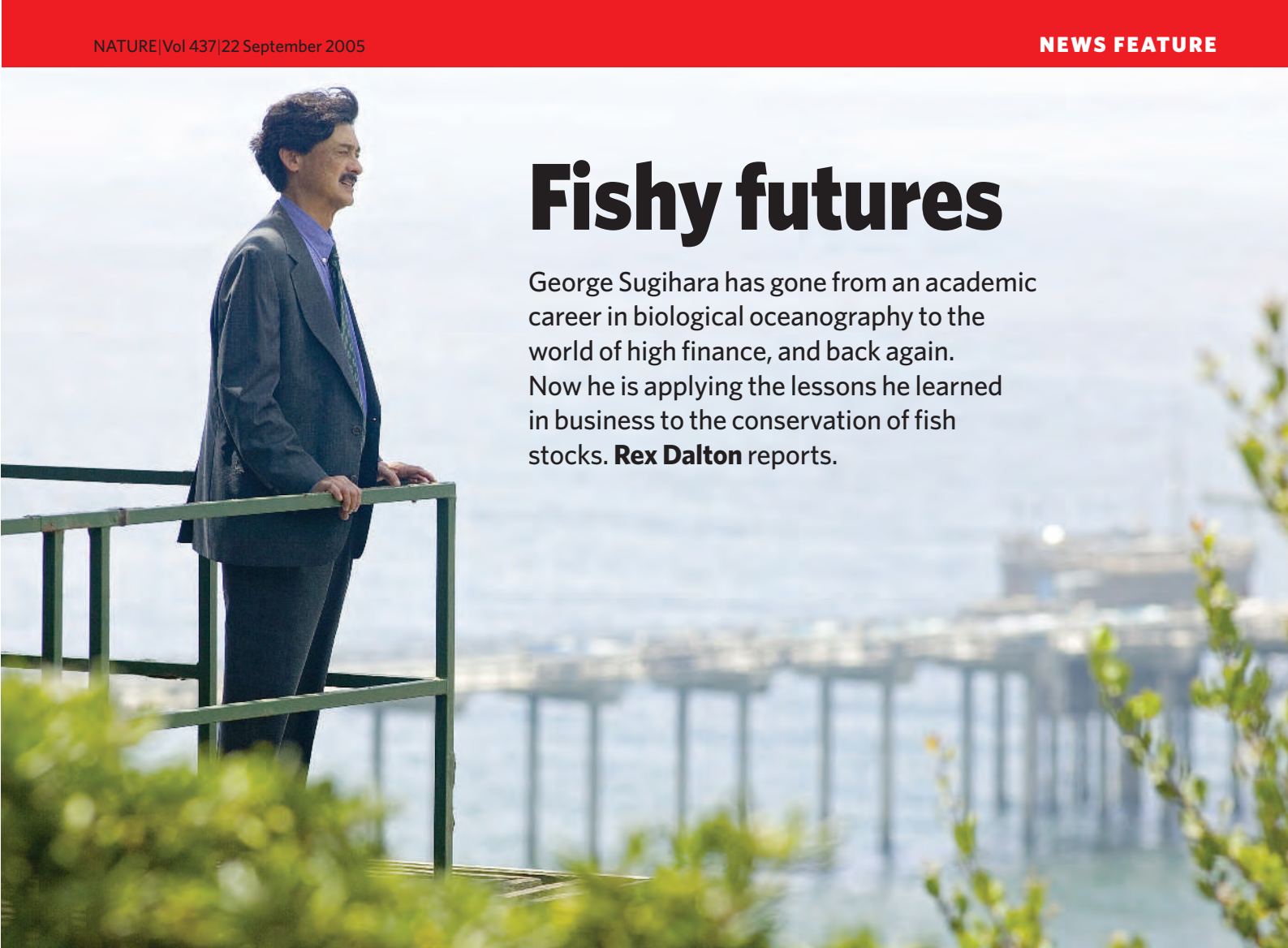
For Ruiz Bravo, it's these research projects that make the K25 programme so valuable. "Life science has moved from involving a single investigator in his or her lab to dealing with bigger types of problems that require large teams," she explains. Although the NIH is also

actively training graduate students in multidisciplinary fields, there are some distinct advantages to recruiting mid-career researchers. "These researchers are already very well trained in their particular area," she says. "They bring a different approach to the life sciences."

That approach is centre stage at the University of Maryland, where Hartley is giving what seems to be an introductory physics lecture. The differential equation on his overhead screen — familiar to any physicist — describes how a gas will diffuse through a room, but Hartley is using it to model how *Chytridiomycota*, a waterborne fungus that causes disease in amphibians, spreads in a population of frogs. His students, mostly epidemiologists, look forlornly at the mathematics overhead, and Hartley is sympathetic. He knows what it's like to be in their shoes: after class ends he will drive 100 kilometres south to attend a three-hour lecture in biostatistics. For now he smiles at his bewildered audience: "You guys wanted to learn more about this physics stuff," he reminds them.

Hartley's students may take more convincing, but his new colleagues are enthusiastic about his changing career. Glenn Morris, who chairs the department of epidemiology and preventive medicine at the University of Maryland School of Medicine, says that he fought hard to bring Hartley to his department. "The physics guys are way ahead of us when it comes to modeling," he says. "For us to make progress, we've got to recognize the importance of these people from alien disciplines."

Geoff Brumfiel is the physical sciences correspondent for Nature in Washington DC.



Fishy futures

George Sugihara has gone from an academic career in biological oceanography to the world of high finance, and back again. Now he is applying the lessons he learned in business to the conservation of fish stocks. **Rex Dalton** reports.

When George Sugihara was seduced by German bankers, he was wooed with style. He was whisked from his San Diego lab to a swanky London hotel, where his room came with a personal butler. Then it was on to a country estate stocked with the finest food and wines. His suitor was a senior executive with Deutsche Bank, one of the world's leading financial firms. One night, with a Cuban cigar in hand, the banker wrote ever-increasing monetary offers on a napkin. "It was remarkable," Sugihara recalls.

By the mid-1990s, banks and investment houses had realized that academics skilled in mathematical modelling could help them to devise winning strategies with which to play the world's financial markets. Sugihara, who had built a formidable reputation among ecologists by analysing the population dynamics of fish and plankton, was a prize catch. Deutsche Bank wanted him to apply those talents to its 'black-box project', a secret endeavour designed to predict the prices of various financial instruments.

Sugihara struck a hard bargain. In addition to providing an ample salary, Deutsche Bank agreed to let him stay in San Diego — where the Frankfurt-based firm provided a posh

office complex overlooking the harbour. There, it gave him all the resources he needed to devise models to decipher price trends from masses of financial data.

In 1995, when Sugihara took leave of absence from the University of California, San Diego (UCSD), his colleagues thought it unlikely that he would ever return — few scientists who have been seduced by the world of finance have later resumed their academic careers. But Sugihara has bucked that trend, and is now applying his experience in finance to marine conservation. He wants to harness market forces to prevent over-fishing — which governments and the scientists who advise them have mostly so far failed to achieve.

In reality, Sugihara never turned his back on biological oceanography. During his four years with Deutsche Bank, he taught part-time at UCSD, and published more than a dozen scientific articles on complex biological systems. When his leave period was up, he says, hard science was always going to win over high

finance. "No, it wasn't hard to leave that world," he says. "I really wanted to do science."

But Sugihara's experience of the markets has changed the way he thinks about managing the ocean's natural resources. For decades, investors have traded on markets for the future prices of virtually every commodity, from grain crops, through orange juice, to oil. Yet despite worldwide sales of at least US\$80 billion a year, there is no futures market for fish. Sugihara hopes to change that. By providing people with the means to make money, and offering a structured financial environment for the worldwide catch and sale of fish, he argues, it should be possible to prevent stock depletion.

Trading places

To this end, Sugihara and a number of scientific colleagues are now seeking start-up finance for a company called the Ocean Resource Exchange. This would trade and lease financial instruments or derivatives associated with fish catches, on an electronic commodities exchange.

Perhaps trading is in Sugihara's genes. His Japanese father was a trader in wood products, who settled in California in 1951 with his Indonesian wife and young son, seeking new opportunities away from the turmoil of

"Basically, I modelled the fear and greed of mobs that trade."
— George Sugihara

G. PAYNE

post-war Asia. But the young Sugihara didn't follow his father into business. After graduating from the University of Michigan in 1973, he embarked on an academic career, initially studying lake cores in Africa. First he worked in Zambia, where he identified pollens and diatoms for palaeoclimate studies. Later, he moved to Tunisia to study algal productivity and the origins of hydrogen sulphide emissions from Lake Tunis.

Sugihara's analytical mind found this fieldwork unsatisfying, so he returned to Michigan to bone up on mathematics. "I took 26 courses in two years," he says. And with his growing mathematical sophistication, he developed a theory to explain an observed regularity in the distribution of species abundance¹. When he approached Robert May, then conducting pioneering analyses of biodiversity at the Institute for Advanced Study in Princeton, New Jersey, with the theory, May immediately recognized Sugihara's potential — and signed him up as a doctoral student.

By the time Sugihara completed his PhD in 1982, he already had his eyes on UCSD's Scripps Institution of Oceanography, which hosted a largely untapped repository of oceanographic and fisheries data. "This was a gold mine," says Sugihara. "And no one was looking at it intensively."

At Scripps, Sugihara used these data to develop and test mathematical models designed to probe the dynamics of complex biological systems. Among the results was an influential article published with May, which showed how to use nonlinear equations — formulas where output isn't proportional to input — to make short-term predictions about the behaviour of chaotic systems such as the population dynamics of marine plankton².

Trend setter

Among those who recognized the equations' power was former behavioural ecologist Steven Schulman, who knew Sugihara from Princeton. By 1990, Schulman was in the New York office of the financial firm Merrill Lynch, conducting quantitative analyses to reduce investment risk. In Sugihara's equations, Schulman saw the possibility of predicting prices in market derivatives. So he brokered a consulting deal: Merrill Lynch provided Sugihara with financial data, which he mined for price trends.

For Sugihara, it was a dream. First, the arrangement allowed him to put his own finances on a sounder footing. "I couldn't afford to send my children to college, back then," he says. Analysing the markets also presented him with fresh intellectual challenges. "I'm driven by access to data," he says.

Then, in 1995, came the extravagant courtship by Robert Stein, then the head of Deutsche Bank's Japanese office. Confidentiality agree-



Nice work: the Deutsche Bank installed George Sugihara in plush offices on the 20th floor of the Emerald Plaza in San Diego.

ments prevent Sugihara from discussing details of his work for the bank. "Basically, I modelled the fear and greed of mobs that trade," he says. And at the time, Sugihara was even more discreet, telling acquaintances who asked about his work: "I'm a teacher." Former colleagues who visited didn't know what to make of his new life as a financial sleuth. Sugihara recalls the first time that May dropped by at his harbour-side office and assumed he was the victim of an elaborate practical joke. "He opened a desk drawer to look for something with my name on it," Sugihara says.

"The motive here is public service. I think we can use market forces for conservation."

— George Sugihara

Sugihara's earnings in the world of finance have provided a home with an enviable sea view, plus a vintage Porsche parked in the garage. But by the standards of banking high-fliers, these are limited extravagances. For Sugihara, acquiring wealth was never the main goal, so he had few qualms about getting back on the treadmill of winning grants for his research. That's not always easy for someone who cuts across disciplines, and whose ideas are often ahead of their time. "It's too far out of the box" is a common comment from reviewers, Sugihara observes.

But unlike his colleagues, whose grant

applications get tossed aside, Sugihara has the luxury of being able to support some of his own research, using a trust fund set up during his Deutsche Bank days. In part, that was how he funded his latest work, an analysis of environmental fluctuations and ecological catastrophes in the North Pacific³. This suggests that fishing quotas may need to be set more conservatively, and adjusted more frequently to compensate for environmental conditions, than is typically the case. "The way fish quotas are set is wrong," says Sugihara. "It doesn't fit nature or reality."

Net gains

The National Marine Fisheries Service (NMFS), which sets quotas in US waters, is at least prepared to listen to this message. When Sugihara gave a lecture in June to a NMFS scientific panel on quota methodology in the North Pacific, his talk went on for two hours — three times as long as scheduled — as agency staff quizzed him on the details. "It was really interesting," says Jeffrey Polovina, a NMFS biological oceanographer who organized the meeting, held in Seattle. "But it was pretty complicated stuff. Most of us don't have the background in chaos theory."

Sugihara hopes that the Ocean Resource Exchange will provide an incentive to preserve fish stocks that doesn't rely on a detailed understanding of complex biological systems, and instead taps into people's baser instincts. "Show them how to make more money," he says. The first derivative is likely to be a futures contract for a certain percentage of a fisherman's catch at an agreed price at a specified time. Another planned derivative is an instrument for trading fish quota allotments, called an 'individual transfer quota'.

"Essentially, these are tradable poker chips or options for fishing rights," Sugihara says. Fishermen and investors could hedge their bets, which should reduce the tendency for catches to swing between boom and bust, and give all stakeholders a tangible financial incentive not to cheat and plunder the ecosystem for the maximum short-term return.

As a test of the idea, Sugihara is modelling the concept using data from a Californian squid fishery — where about 200 vessels bring in a haul worth up to US\$36 million per year. But both catches and prices can fluctuate widely, making it a prime candidate for a market in derivatives. "The motive here is public service," he says. "I think we can use market forces for conservation."

Rex Dalton is Nature's US West Coast correspondent.

1. Sugihara, G. *Am. Nat.* **116**, 770-787 (1980).
2. Sugihara, G. & May, R. M. *Nature* **344**, 734-741 (1990).
3. Hsieh, C.-H., Glaser, S. M., Lucas, A. J. & Sugihara, G. *Nature* **435**, 336-340 (2005).

BUSINESS

Swooping for biotech

Big pharmaceutical companies are moving swiftly to acquire biotechnology companies — especially if they can snap them up on the cheap. Meredith Wadman reports.

When Novartis announced a bid to buy Chiron, the California biotechnology company, earlier this month, the Swiss drugmaker trod a well-worn path. Most 'big pharma' companies know that their future rests in either buying biotechnology companies or getting into bed with them.

But as Chiron's directors rejected Novartis' first offer, analysts were asking whether a takeover would pep up Chiron, which has been hit by recent vaccine-manufacturing woes — or if it would just mark a release of cash to shareholders, and the end of a biotechnology success story dating back to 1981.

"Pharma is struggling on its own," explains Karl Heinz Koch, an analyst who follows Novartis for the Swiss bank Lombard Odier Darier Hentsch. "It needs the more dynamic biotechnology industry to deliver growth and value to investors."

Big drug firms "have lots of cash on their balance sheets", adds Geoffrey Porges, a biotechnology industry analyst with investment-research company Sanford C. Bernstein in New York. "They are now saying: 'We don't see enough opportunities internally to invest that cash, so we're going to look externally.'" Novartis is a case in point: the company saw its profits grow by 15% in 2004, to \$5.8 billion, on sales of \$28.2 billion.

Healthy appetite

But a deal that looks tasty to big pharma may appear less enticing from the other end. "One of biotech's major challenges is to keep itself distinguished from pharma, which has more headaches and crises than you can shake a stick at," says Arthur Caplan, an industry observer and bioethicist at the University of Pennsylvania in Philadelphia. Keeping a distinctive profile is "a real challenge" for the biotechnology sector, Caplan points out.

Novartis' wooing of Chiron is hardly a blind date. The drugmaker already owns 42% of the biotechnology company, inherited from Ciba-Geigy when it merged with Sandoz to create Novartis in 1996. And despite Chiron's coyness, it may welcome the interest. Chiron has had a rocky year since problems at its vaccine plant in Liverpool, UK, were revealed last October. Influenza vaccine sales alone fell by



Drug target: Chiron could be facing an end to 24 years of independence — and research job cuts.

\$178 million in 2004, and profit margins fell by more than a fifth.

On 1 September, Novartis offered \$40 a share, or roughly \$4.5 billion, for Chiron, whose share price rose by 18% to almost \$43 (see graph). Four days later, the California company's directors pronounced that offer "inadequate". Negotiations are continuing, with reports that Novartis may increase its bid.

The putative deal reflects a growing appetite among the major pharmaceutical companies for snapping up small, creative biotechs with their innovative research and potentially lucrative products. GlaxoSmithKline joined in on 7 September, when it announced its agreement to purchase ID Biomedical, a Vancouver, Canada-based vaccine maker, for \$1.4 billion.

Novartis followed up its Chiron bid by announcing a subtler partnership with another biotechnology company. It will buy a 20% share in Alnylam Pharmaceuticals of

Cambridge, Massachusetts, a collaboration aimed at developing drugs based on RNA interference, which uses recombinant DNA to block disease-causing genes.

The two deals may represent opposite ends of the spectrum in terms of how large drug companies treat partners in the biotechnology sector. In the case of Alnylam — founded by Nobel-prizewinning biologist Phillip Sharp of the Massachusetts Institute of Technology — the arrangement envisages Novartis pouring money into the smaller company, whose laboratories would basically retain their autonomy.

CHIRON

Shopping spree

In contrast, some analysts think that a takeover of Chiron would shrink the company's research. This employs more than 380 scientists at laboratories in California, with others elsewhere, and cost \$431 million last year. "There's a lot of cost savings Novartis can take out of Chiron," says Koch. "The large majority of its research and development cost is in biomedical research, which has not produced a single drug in eight years." Analysts also speculate that Novartis may finance any acquisition in large part by selling off Chiron's lucrative blood-testing business.

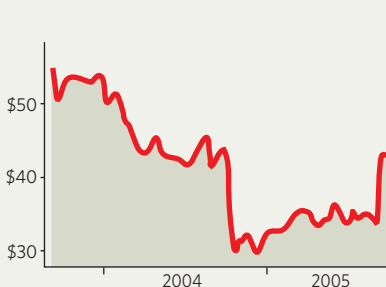
New York-based Pfizer has been even more active in biotechnology acquisitions. This year alone, it has acquired three privately held California companies: Bioren, Idun Pharmaceuticals and Angiosyn. And last week it bought Vicuron, a Pennsylvania-based maker of anti-infective drugs, for \$1.9 billion.

Drug companies have long bought biotech companies as a means of expanding their drug pipelines, but several factors have converged to accelerate the process. The industry has faced bad publicity — most notably surrounding Merck's withdrawal of the painkiller Vioxx (see *Nature* 436, 1070; 2005). And as lucrative drugs go off-patent, drug companies are under pressure from their shareholders, who may be placated by smart acquisitions.

It is also a good time to be shopping. Public offerings for biotechnology companies have generated disappointing returns lately, making companies more willing to be bought out. And stagnant biotechnology share prices make acquisitions relatively cheap. "This hasn't been a great year for biotechnology stocks," notes Brady Huggett, managing editor of *BioWorld*. "So biotechnology companies have often been pretty decent bargains."

Chiron's share price, for example, has slumped from \$57 in late 2003 to less than \$37 when Novartis made its bid. "The problems in the vaccine business have caused Chiron's whole business to be undervalued," says Porges. "Novartis is being opportunistic in trying to buy Chiron when it is back on its heels."

CHIRON STOCK PRICE



Re-wilding: no need for exotics as natives return

SIR — In their Commentary “Re-wilding North America” (*Nature* **436**, 913–914; 2005), Josh Donlan and colleagues propose introducing Asian and African species to the Great Plains. But they do not discuss a real effort that is already under way to restore native North American prairie wildlife on the Northern Great Plains.

The World Wildlife Fund and its partner, the American Prairie Foundation, have launched an ambitious programme to purchase, from willing sellers, property in north-central Montana. When combined with adjacent public lands, this would provide the habitat for nearly the entire suite of North American grassland species that have lived here within the past 10,000 years.

These efforts envisage reintroducing bison from remaining genetically pure herds and providing habitat that will support increasing populations of nearly extinct species such as black-footed ferrets. If all goes well, there will be an increase in populations of pronghorns, elks, mountain plovers, burrowing owls and large predators such as mountain lions — our native felid, which has already recolonized this area without human intervention. At least parts of the megafauna-dominated landscape can be restored in a few decades, and Pleistocene survivors such as bison will once again be able to play their role as ecological engineers.

This restoration of native prairie wildlife is being carried out in cooperation with local landowners and communities. It addresses concerns about the return of prairie species that, in some cases, have been absent for a century or more. Restoring the native fauna of this region first is a more economically viable and ecologically sound approach, if the goal is to energize positive support for conservation in general.

Eric Dinerstein*, W. Robert Irvin†

*Conservation Science, World Wildlife Fund,

†US Ecoregional Conservation, World Wildlife Fund, 1250 Twenty-Fourth Street, NW, Washington DC 20036, USA

Re-wilding: don't overlook humans living on the plains

SIR — Proposals made by Josh Donlan and colleagues to “re-wild” the Great Plains (“Re-wilding North America” *Nature* **436**, 913–914; 2005) assume that if the land is void of people, it is necessarily open to exotic megafauna. As a historian of the twentieth-century American West, I disagree, and I believe the re-wilding plan would be harmful to current environmental efforts in the area.

The human population may be sparse, but

people on the plains use large areas of land to drive the economies of the towns that dot the landscape. In the late 1980s, a group of well-meaning people tried to gather support for the Big Open project. This was part of a larger proposal, called the Buffalo Commons, to establish a huge preserve for bison covering 139,000 square miles in ten states from Texas to Montana. Local people overwhelmingly rejected the proposal. Subsequent anti-environmentalist and anti-government feeling damaged efforts that were being made towards environmental sustainability.

But local alliances can be productive, and the stubborn search for middle ground has led to some recent victories for biodiversity in the region. Bison have been reintroduced to Native American reservation lands, land has been restored by Nature Conservancy, plans are in progress to pay ranchers to reduce the number of cattle grazing, and coal bed methane pollution has been opposed by the Northern Plains Resource Council. Some ranchers have taken up environmentally friendlier practices, such as adjusting cattle grazing on the Plains, by use of fencing, to mimic the habits of bison: intensive grazing for a shorter period of time.

Politicians, ranchers and academics have started talking to each other constructively. Can we honestly now ask the region to ingest lions and cheetahs?

Steven Shay

Department of History, Washington State University, Pullman, Washington 99164, USA

How synthetic biology can avoid GMO-style conflicts

SIR — Your News story “Synthetic biologists face up to security issues” (*Nature* **436**, 894–895; 2005), defines synthetic biology as the ability “to create complete genomes from scratch and to introduce new characteristics into viruses and bacteria”. But the second half of this definition has already been applied for decades to genetically modified organisms (GMOs), and particularly to modified viral genomes. The present discussion about regulation of synthetic biology should carefully consider how and why GMOs are regulated, in order to avoid regulatory chaos.

The US and Canadian systems for GMO regulation are based on the properties of the organisms produced, whereas the European system is based more on techniques. The incompatibility between product-based and technique-based systems is the source of much of the transatlantic tension regarding GMOs.

North American scientists are calling for technique-based regulation of synthetic biology. But for products of synthetic biology that bear novel genes and thus are also GMOs, which type of regulation should prevail: technique- or product-based? If the former,

one would quickly encounter the situation where equivalent organisms, synthetic or classic GM, would be regulated using drastically different strategies and criteria. If the latter, the most potentially dangerous products of synthetic biology would simply be regulated as GMOs. If the United States and/or Canada go forward with technique-based regulation of synthetic biology, a minimum of coherence would require them also to shift to technique-based regulation of GMOs — a major policy change.

I believe that the first step to reassure the public about synthetic biology should be to cool the rhetoric. The present situation is reminiscent of 30 years ago, when some of the pioneers in the then-new field of genetic engineering made unrealistic claims about what was feasible; this was one of the major early sources of public uneasiness about GMOs. There should be a bit more modesty in claims both about what can be achieved by synthetic biology in the foreseeable future, and about what could be achieved by additional regulatory supervision.

Mark Tepfer

ICGEB Biosafety Outstation, Via Piovega 23, 31056 Ca' Tron di Roncade, Italy

Chiropractors start major study of spinal outcomes

SIR — Your News story “Survey questions safety of alternative medicine” (*Nature* **436**, 898; 2005) quotes Edzard Ernst as saying that complementary and alternative medicine (CAM) organizations are not doing enough to monitor adverse reactions. It also reports that chiropractic treatments sometimes have serious side-effects.

The British Chiropractic Association (BCA) is currently undertaking a large-scale observational study (sample size of over 50,000) to document patient outcomes after cervical spine manipulation. Final data analysis is expected in 2006, and we hope to publish the results in peer-reviewed journals.

The BCA has also, in conjunction with the Anglo-European Chiropractic College, set up a chiropractic reporting and learning system; more than 1,200 practitioners who are members of the BCA have recently received an information pack to enable them to participate in the scheme. Resulting data will be analysed at the Anglo-European Chiropractic College and outcomes will be relayed to the profession, through our newsletter, journal and website, so practitioners may learn from the experience of others.

The intention is that the scheme will, if successful, be offered to other chiropractic associations within Europe in 2006.

Barry Lewis

British Chiropractic Association, Blagrove House, 17 Blagrove Street, Reading RG1 1QB, UK

COMMENTARY

A universal register for animal names

Andrew Polaszek and colleagues propose an open-access web-register for animal names, which they believe is vital to move taxonomy into the twenty-first century.

How can we maintain and continue to benefit from our planet's biodiversity? A first step, the effective exchange of information about biodiversity, needs an efficient and stable means of naming species. For animals, this is achieved with the Linnaean system of binominal nomenclature, introduced in 1758, and a comprehensive set of rules administered by the International Commission on Zoological Nomenclature (ICZN). Although the Linnaean system and the ICZN code have been hugely successful, they are often perceived as failing to meet the needs of today's biologists. To meet these new demands, we propose the creation of a mandatory web register for all new animal names, and the subsequent inclusion of all existing animal names and nomenclature in a single information system.

With more than 1.5 million animal species described so far, the scale of taxonomic study today poses considerable challenges to a system developed in the age of letter-post and printing press — when the community of taxonomists and biologists was much smaller. Each year 15,000 to 20,000 new animals are named according to the rules of the current code (ICZN, 4th edn, 1999). These new names and descriptions are scattered across many journals and other publications. In entomology alone, taxonomically relevant information can be found in more than 1,100 specialized journals.

Moreover, sources such as books and conference proceedings are difficult to access or have low print runs. Where the code allows publication in different types of media, this adds to the complexity. These problems not only affect the progress of taxonomy, they also make it harder for taxonomy to be used by non-specialists.

One for all

We propose a register of new zoological names — ZooBank — to be established and administered by the ICZN, and bolstered by a mandatory requirement, in the next edition of the code, for the registration of new names. The register would be web-based and open-access, and would cover all taxonomic ranks relevant to the code.

The idea of a register for newly discovered organisms is not new. Such a register already

exists for bacteria, and was considered and rejected by the plant-taxonomy community who decided that with fewer taxa (and some excellent existing databases) it was not needed.

We believe that ZooBank could have huge benefits for taxonomists and for biologists in general. First, many names currently published do not conform to the current code, and sorting out the ensuing mess wastes time. ZooBank would improve code-compliance by using automated tools that are integrated into the registration process. We stress that assessing the merits of different taxonomic hypotheses would not be part of ZooBank's function; it would be a register, not a peer-evaluation system. Second, anyone could sign up for automatic e-mail alerts, which would notify them of any additions or changes to the taxonomy of a particular group. Third, it would democratize taxonomy and allow those without access to major libraries to retrieve essential information.

The current code is a complex document, and there are many technical issues involved in developing an online register. These include establishing the precise date when a name becomes available (at publication or registration), effective feedback systems for correcting errors, and the problem of archiving. We are confident that these issues can be successfully addressed, as they have been for molecular databases such as GenBank.

Eventually, ZooBank could allow retrospective registration of existing names, and of all nomenclatural acts in zoology. Although this would require considerable resources, ongoing projects that seek to catalogue existing names should help immeasurably. Support from Zoological Record — the closest thing currently available to a register of animal names (www.biosis.org.uk/ion) — is particularly important. Other resources are the uBio nameserver (www.ubio.org/nameserver), Species 2000 (www.sp2000.org), Integrated Taxonomic Information System (www.itis.usda.gov) and GBIF's Electronic Catalogue of the Names of Known Organisms (www.gbif.org/prog/ecat). Rather than replace these projects, ZooBank will link to

them and provide the definitive naming authority.

Joining forces

Through its website (www.iczn.org), the ICZN is initiating a year-long period of consultation on the merits of mandatory registration and the details of ZooBank's creation. Using existing resources, the register will be established and will accept names on a voluntary basis. Compulsory registration will begin only with community support and when resources to run the project for at least ten years are in place (with the expectation that it would continue indefinitely). Our target is to have ZooBank, and a fifth edition of the code, completed by 2008, the 250th anniversary of Linnaeus' animal nomenclature.

Taxonomists are often criticized for failing to act together as a community, not least in *Nature*. However, the almost universal, voluntary adherence to the current codes of nomenclature is arguably one of the strongest examples of international scientific cooperation. The success of the ICZN in facilitating this cooperation over many years makes it the right organization to spearhead a universal system for the registration of zoological names. We appeal to all taxonomists to support this project and to engage in the consultation needed to design the best system. We also appeal to all biologists, whose work depends on taxonomy, to throw their weight and influence behind this initiative.

It is inevitable, and to be welcomed, that taxonomy will rely increasingly on electronic forms of communication. Molecular methods in taxonomy, such as the current Barcode of Life initiative, will increase in importance. Integrating ZooBank with such projects will be critical in maintaining the coherence of taxonomy and avoiding conflicting systems of names. What we propose will make animal taxonomy a truly modern science. ■

Andrew Polaszek is Executive Secretary of the International Commission on Zoological Nomenclature, based at the Natural History Museum, London.

The list of co-authors can be found on *Nature's* website as Supplementary Information (www.nature.com/nature).

"We appeal to all biologists whose work depends on taxonomy to throw their weight behind this initiative."

BOOKS & ARTS

Coping with interesting times

Unless we take action urgently, climate change could spell disaster for a wide range of living species.

Climate Change and Biodiversity

edited by Thomas E. Lovejoy & Lee Hannah

Yale University Press: 2005. 440 pp. \$65, £45

Paul Colinvaux

A Chinese curse says: "May you live in interesting times." The 'interesting times' the Chinese had in mind were familiar stories from human history: battle, murder and sudden death, chaos, disease and famine. But the authors corralled by Tom Lovejoy and Lee Hannah talk of times even more 'interesting' than these: they discuss the grotesque climate changes of the greenhouse Earth. These are interesting enough for the chances of human happiness but are truly desperate for the survival of the remaining diversity of life on Earth. In the 24 chapters of *Climate Change and Biodiversity*, 66 authors suggest a little of what might be done to save something from the wreck. The book is written in the hope that policy-makers might read it, and it has been skilfully edited to that end by turning as much scientific prose as possible into plain English.

The brutal outline of what is happening is a twice-told tale. The atmosphere and continents are warming as we release carbon dioxide into the air through our habit of powering our society by oxidizing fossil carbon. Some of the first, most easily predictable, results are already apparent as the Arctic permafrost and glaciers melt. A rising concentration of carbon dioxide, a rising temperature and a consequent rising sea level have all been measured. These early-warning signs of what's to come are well known and have their own literature, so Lovejoy and Hannah offer just a few chapters on climate and modelling for those who need to play catch-up. Without profound changes in our energy systems, we are in for a few generations of misery, but as with other 'interesting times' we shall muddle through, forced to cope with the conditions. It is the rest of life on Earth that concerns the authors.

Climate is part of geography. The air is divided into great masses, much as the land is divided into continents, although the two maps are not congruent. But maps of the usual or seasonal positions of air masses over continents and maps of vegetation are congruent. Indeed, they are so congruent that the first climate maps were made by mapping vegetation on a continental scale and calling the result

'climate'. Great patches of plants, large enough to be coloured in the family atlas, were called 'formations'. Later, ecologists complained that animals were being neglected and were probably as typical to the climate as were the plants. They named a terrestrial unit with characteristic climate a 'biome', a name that has stuck, even though it is a mere abstraction.

A conservationists' lament echoed in this book is that most nature reserves and conservation areas are now in the wrong places. Some, such as Yellowstone National Park, were set up on land that was little good for anything else; others were created to protect a good bit of a biome, or a place where the cuddly or the rare were known to live. But the greenhouse is moving the air masses, so the old places of refuge will be no use. What is to be done?

Strangely, palaeoecology comes to our rescue. We know from fossils, particularly from fossil pollen, how the plants of the great Pleistocene formations survived the repeated redrawing of the climate map of the ice-age Earth. Plants left the formations of their ancestors and dispersed to new habitats, or hung on in patches where the new climate was within the bounds of tolerance, or coexisted at the edges of old ranges with plant survivors of what was once an adjacent range.

Plants move as seeds, carried by winds or animals, whereas animals can move on their

own. But they can all move, and they all do: none of them are forced to remain in what used to be a refuge or a relationship. The answer to the conservationist dilemma, then, is to let every refuge have an escape route. Conceive a world of scattered refuges in between patchworks of land parcels of mixed use: built-up areas, parks, houses with gardens, green strips on highways, windbreaks. Conservationists call this design the 'matrix' — a meaning in keeping with the word's derivation from the Latin for 'womb'. If we concentrate on providing a good matrix between conservation areas, we just might solve the problem.

The politics will be difficult. Some of the book's authors pin their hopes on trying to stop us building the greenhouse. My own cynical view is that if the government of a great nation will not act to protect the livelihoods of its own citizens, it is unlikely to take political risks to defend a few animals and plants. Lovejoy and Hannah's book may well be the only kind of resource we have: an attempt to let those who have the power to act know what the intelligence system of science is saying, and that it should be believed. This is where the battle must be fought. Now we need some even more readable essays.

Paul Colinvaux is in the Marine Biological Laboratory, 7 MBL Street, Woods Hole, Massachusetts 02543, USA.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Pack your trunk: wildlife
will need to be mobile to
cope with climate change.

A path through the forest

Nerve Endings: The Discovery of the Synapse

by Richard Rapport

Norton: 2005. 224 pp. \$23.95

Jeffrey S. Isaacson

It is not often that one can pinpoint a paradigm-shifting moment of discovery that transforms a scientific discipline, but in the 1800s, one such moment erupted from a makeshift laboratory in the kitchen of a young Spanish professor. Using not much more than a simple light microscope, pen and paper, and specimens of brain tissue, Santiago Ramón y Cajal (1852–1934) formulated revolutionary concepts that ignited the field of neuroscience.

Nerve Endings, a book from neurosurgeon Richard Rapport, recounts Cajal's life and times. The book also sheds light on Italian histologist Camillo Golgi (1843–1926), whose own kitchen experiments led to the tissue-staining method Cajal used to great effect. But even though both men's accomplishments are intertwined, and led to their joint winning of the Nobel prize in 1906, Cajal and Golgi were scientific rivals with opposing views on the nature of the brain.

Early microscopic studies hinted that the nervous system was made of individual cells (neurons), with numerous protrusions (dendrites) and single, thin emanations (axons). It was another matter to imagine how neurons were organized to convey information.

A popular view in the nineteenth century was that neurons were continuous with one

another and formed one gigantic network. This 'reticular theory' seemed quite reasonable as it was an ideal way for information such as sensory input and motor output to flow through the nervous system in both directions. But inefficient techniques for fixing and staining tissue plagued the study of brain microstructure. Typical procedures labelled virtually all neurons and fibres, revealing the forest, but hiding the individual trees, leaves and roots in the thicket.

Golgi, a reserved figure, overcame this dilemma by developing an almost alchemical process based on soaking brain pieces in potassium dichromate and silver nitrate. The result was striking — only a tiny fraction of neurons were impregnated with a dark silver precipitate, making it possible to follow the outline of a single neuron and its tiniest processes. Why this method, which Golgi named the black reaction, labels neurons with such exquisite randomness remains unknown.

Golgi made important observations with his new technique, including that single axons give rise to numerous branches (collaterals). The fine meshwork of axon collaterals represented to Golgi the reticular elements that linked neurons, whereas dendrites played merely a nutritive role.

Much is known about Cajal's life from his eloquent autobiographical writing. His *Recollections of My Life* (MIT, 1989) should be required reading for any scientist and Rapport draws heavily from it. The son of a small town doctor, Cajal was a mediocre student, obsessed with art, bodybuilding(!) and chess. His compulsive mind became intrigued with deciphering the workings of the brain. Working alone with few resources and only books to guide him, Cajal's neuroscience was done with a passion and fury unrivalled to this day. To him, "an exact knowledge of the structure of the brain was of supreme interest for the building up of a rational psychology. To know the brain ... is equivalent to ascertaining the material course of thought and will."

In 1887, Cajal was introduced to the Golgi staining method and immediately grasped its power. Whereas most neuroanatomists studied adult human brains, Cajal focused on embryonic tissue and the more compact brains of mice and birds. Gazing into his microscope, Cajal conjured "the new truth": axons ended and formed contacts (later termed synapses by Charles Sherrington) very close to the dendrites and cell bodies of other neurons. To Cajal, the reticular theory was



Kitchen think: Santiago Ramón y Cajal in his home lab.

clearly wrong and the 'neuron doctrine' was set to take its place. He further proposed the law of 'dynamic polarization' which stated that neurons received information at their dendrites and cell bodies and relayed nerve impulses through their axons.

Cajal saw that by studying the relationship between axons and dendrites he could infer in what direction information travelled across neural circuits. His meticulous drawings of brain circuits — works of science and art — are filled with playful arrows showing the direction of information flow. Frustrated by the slow pace of publishing, in 1888 he created his own journal to showcase his findings. When that failed to make an impact, the following year Cajal travelled to Berlin to demonstrate his slides at a meeting of leading anatomists. Fascination over Cajal's preparations led to the rapid confirmation of his conclusions. Despite the quick acceptance of Cajal's neuron doctrine, Golgi stubbornly defended his ideas on reticular theory. It is ironic that the two met for the first time in Stockholm to receive their prizes; Golgi's speech was an attack on neuron theory, Cajal followed with a tactful defense.

As a synaptic physiologist a century later, I find it astonishing that Cajal's simple observa-

NEW IN PAPERBACK

Linnaeus' *Philosophia Botanica*

translated by Stephen Freer

(Oxford University Press £35, \$89.50)

The first full English translation since 1775 of this classic book summarizing Linnaeus' work on the classification and taxonomy of plants

Uncertain Science ... Uncertain World

By Henry Pollack (Cambridge University Press £12.99, \$19.99)

Human-Built World: How to Think About Technology and Culture

by Thomas P. Hughes (University of Chicago Press, \$13, £9.50)

"A virtuoso overview of the various relationships between technology, commerce, society, art and the military." *Graham Farmelo Nature* **429**, 348 (2004).

Subtle is the Lord

By Abraham Pais (Oxford University Press £17.99, \$22)

A new edition, with a foreword by Roger Penrose, of this masterly biography of Albert Einstein.

tions led him to infer not only that synaptic contacts existed, but also that the unseen synaptic gaps were the basis of neurotransmission. With uncanny accuracy, he foretold today's hottest research: Cajal envisioned that tiny dendritic protuberances (spines) were important postsynaptic elements and that learning and memory might reflect the growth of new contacts. Rapport does well in making

Cajal accessible to non-neuroscientists. And for those of us who still ponder his drawings to guide us through the tangled forest, Rapport reminds us how lucky we are to have Cajal as our guide and companion. ■

Jeffrey S. Isaacson is in the Department of Neuroscience, University of California, San Diego School of Medicine, La Jolla, California 92093-0608, USA.

examples of the phenomenon. In the modern age, however, patients no longer have the final say on whether they are sick, as modern medicine tends to rely on the signs its techniques and technology can detect — it can even spot hypertension or latent cases of hepatitis C in asymptomatic people. Alternatively, unable to detect the biological underpinnings of conditions such as chronic fatigue syndrome, the profession uses its cultural authority to cast doubt on the reality of the complaint.

Love has symptoms too, for it has profound physical and psychological effects. Duffin points out that in an era of symptom-based medicine, 'lovesickness' emerged as a clinically respectable disease characterized by such symptoms as anorexia, insomnia and melancholy. It turns out to have an ancient pedigree and lingered in the repertoire of diseases until remarkably recently. Duffin even argues that love "still carries disease overtones in the medical and cultural psyche" today, something she attributes to the fact that it threatens to bring the "loss of control".

I found this portion of the book only intermittently compelling and occasionally self-indulgent. By contrast, Duffin's final chapter on the rise of hepatitis C fizzles with information and ideas that draw the reader into her argument. No one who reads this book will ever again think there is a 'routine' blood transfusion, for example, the hidden perils of this often life-saving procedure being starkly laid bare. More broadly, Duffin's examination of the proliferating array of hepatic diseases leads one to reflect upon a series of ironies surrounding the diagnosis and treatment of disease in the modern era, and to confront the continuing intrusion of moral values into the supposedly value-free realm of medicine.

Consider, for example, the construction of 'guilty' and 'innocent' victims of AIDS and hepatitis C; or the legal and ethical morass that surrounds the question of whether those infected by the blood supply should be financially compensated for their suffering, and by whom; or the complicated and to some degree perverse consequences for individual patients of the discovery that they have a symptomless, untreatable disease that may or may not lead to debility and death at some unknowable time in the future. Life lived beneath the sword of Damocles acquires a whole new meaning, and not one that most of us would welcome. Although as Duffin would be the first to point out, in reality we all exist in such a state anyway, and contrive to hide this painful truth from ourselves at all costs. ■

Andrew Scull is in the Department of Sociology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0533, USA.

An ill-defined idea?

Lovers and Livers: Disease Concepts in History

by Jacalyn Duffin

University of Toronto Press: 2005. 240 pp. \$55, £35 (hbk); \$27.50, £20 (pbk)

Andrew Scull

Jacalyn Duffin's brief but beguiling book (it contains only 127 pages of text, the rest being consumed by notes and bibliography) is a revised version of the Joanne Goodman lectures, which she delivered at the University of Western Ontario in London, Canada, in 2002. Judging by the sprightliness of her prose, she must have provided an entertaining time for her audience. Along the way, her listeners will have encountered a clever series of arguments for viewing diseases as ideas, and a sometimes passionate dissection of disease and illness, doctor and patient, culture and pathology. At her best, Duffin creates a genuine sense of excitement and engagement with her materials, and these qualities are nowhere more evident than in her concluding chapter on livers (diseases thereof), where she draws fruitfully on her own clinical experience as a haematologist.

The notion that diseases may be thought of as "ideas influenced by the tastes and preoccupations of society" is likely at first to raise hackles in some quarters. Not another postmodernist rant, some will sigh, and sure enough, on the very first page there is a reference to the archfiend Michel Foucault. Not to worry, Duffin is not in the business of denying biological realities. Her point is far more clever and subtle, revolving around a distinction that she is scarcely the first to draw between disease and illness. This is, she concedes, a linguistic convention that will never be observed at the level of everyday speech, but making it here allows us to talk intelligently about two very different phenomena. Illness "applies to the subjective aspects of suffering, the problem experienced by the individual patients" and changes very

slowly, if at all, over time; disease, in contrast, refers to "our ideas about that illness", ideas that not only describe symptoms and sufferers, but also incorporate an explanation or a theory about the illness. And diseases often change and proliferate as we recognize and add entities to the list of ailments that afflict the human race. It is this process of medicalization and demedicalization that Duffin's substantive chapters seek to illuminate.

Illnesses — that is, the symptoms experienced and reported by patients — often loom large in the making of a disease concept. Indeed, Duffin contends, "During early modern times, one could not be sick without feeling sick," and even today there are diseases still constructed solely on the basis of symptoms. The 'mental illnesses' that fill the pages of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* are the most numerous, although scarcely the only,



Call a doctor: when diseases were defined by symptoms alone, lovesickness was a clinically respectable diagnosis.

An anatomist among the artists

William Hunter at the Royal Academy of Arts.

R. COLL PHYSICIANS



Show 'em how it's done: physician William Hunter explains the details of the human form to an audience of artists. He also used his position to lecture on the relationship between art and nature.

Martin Kemp

When the newly founded Royal Academy of Arts appointed William Hunter, the Queen's obstetrician, to be their first Professor of Anatomy in 1769, they were following European precedent, albeit belatedly. From the sixteenth century onwards, academies of fine arts, dedicated to elevating art from craft to intellectual pursuit, had taught anatomy, perspective and other branches of knowledge deemed essential for the 'learned' painter or sculptor.

No doubt the academy expected its new professor to 'stick to his last' — a phrase derived from Pliny's anecdote about Apelles' reply to a shoemaker. When a cobbler criticized the great Greek painter for mistakes in his portrayal of the sandal, Apelles was happy to listen, but when the artisan went on to criticize the painting of a leg, the painter's riposte was blunt. Hunter's specialist 'last' would have been to instruct aspiring masters in the structures of the human body, above all the bones and muscles that determined posture and expression.

In the event, the academy got more than it bargained for. In his series of lectures, preserved in incomplete notes in the University of Glasgow, Hunter ranged far beyond the ostensible subject of his teaching. He volunteered firm views on that

most vexed of subjects, the 'imitation' of nature. His view, as a man of British eighteenth-century science committed to an unrelenting empiricism, was that the role of art was to achieve as exact an imitation of natural appearance as possible.

He declared that "the superiority of Nature over Art seems to shine forth in almost every thing". Thus, "in the Fine Arts the more precise the imitation of Nature is ... the more striking I should suppose the effect will be". He concludes that "a painter or sculptor ... cannot copy Nature too exactly, or make deception too strong". His models of perfect imitation were the coloured casts of wombs and foetuses that he had made in conjunction with his great 1774 book of obstetrics, *The Anatomy of the Human Gravid Uterus*.

The ambitious president of the new Academy, Joshua Reynolds, would not have been able to stomach such a radical view of naturalistic imitation. He was committed to the idealizing aesthetic traditionally promulgated by the academies. He told his audience in his own lectures that "nature herself is not to be too closely copied", as "a mere copier of nature can never produce anything great". Reynolds's ideal was to imitate the great masters of antiquity, and their peers, such as Raphael and Michelangelo.

The presence of Reynolds at one of

Hunter's lectures is recorded in a fascinating painting by Johann Zoffany (above), now on show until 27 November at the National Portrait Gallery in Edinburgh, in an exhibition celebrating 500 years of the Royal College of Surgeons of Edinburgh.

Hunter is actively engaged in demonstrating anatomical points from a living model. He is also accompanied by a suspended skeleton and a life-size cast of a flayed criminal. Reynolds, along with fellow academicians and a few token students, listens to the good doctor through his ear trumpet.

Reynolds might have marginalized Hunter's views, dismissing them as coming from someone who was not a professional authority on the Fine Arts (even if he was a significant collector). But the future was to lie more with Hunter's direct and raw access to nature than with the president's lofty idealism. The naturalistic tendency, epitomized in Britain by John Constable's landscapes, was to become a rising force in European art, redefining the 'science of art' in the direction of empiricism rather than the Platonic ideals espoused by the academicians. Hunter's insistent voice clearly played a role in this process.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK.

NEWS & VIEWS

ENVIRONMENTAL SCIENCE

The carbon cycle under stress

Dennis Baldocchi

In the summer of 2003, Europe experienced an exceptionally hot and dry spell. That 'natural experiment' prompted a continental-scale analysis of how terrestrial ecosystems respond to such climatic extremes.

Plant ecosystems are major players in the carbon cycle — they take up carbon dioxide in photosynthesis and release it in respiration at rates that are nonlinear functions of temperature. How net carbon exchange by terrestrial ecosystems will be affected by projected global warming, and by the increased incidence of extreme climatic events, is a question under intense investigation. Ciais *et al.*¹ (page 529 of this issue) have taken advantage of an episode of heat and drought that affected Europe during the summer of 2003 to evaluate how net carbon exchange of ecosystems responded.

The context to this research is well known. Anthropogenic combustion of fossil fuels has caused mean concentrations of CO₂ in the atmosphere to reach and exceed 380 parts per million (p.p.m.)², a level that is about 100 p.p.m. greater than in pre-industrial times³. Because CO₂ absorbs long-wave energy, it warms the Earth's atmosphere. Additional inputs of carbon to the atmosphere will produce further warming and may contribute to the occurrence of more-intense spells of heat⁴. But how can we estimate the effects of warming trends and extreme events on ecosystem photosynthesis and respiration?

Manipulative experiments that warm ecosystems with infrared lamps provide some insight into these questions⁵. However, such studies are generally confined to short vegetation in small plots, so their findings do not scale well to the dimensions of forests and continents. Natural experiments are an alternative approach to studying how ecosystems may respond to warming. They also provide information on how ecosystems co-vary with elevated CO₂ and ozone, and with a reduction in rainfall. But such experiments are feasible only when environmental conditions exceed some threshold or range of conditions. For much of the past century, natural warming experiments have been hindered by the fact that many potential cases have fallen within the natural climatic variability attributed to events such as El Niño/La Niña, volcanic eruptions, the North Atlantic Oscillation and the Pacific Decadal Oscillation⁶. This situation is now changing as global and regional temperature trends begin to rise above natural climatic variability⁷.

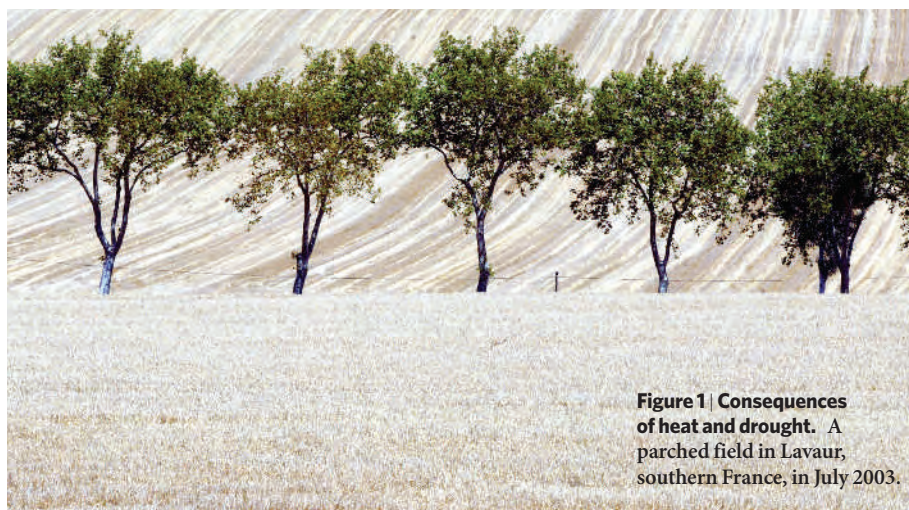


Figure 1 | Consequences of heat and drought. A parched field in Lavour, southern France, in July 2003.

The summer of 2003 in Europe is a particular case worth studying (Fig. 1). Rainfall dropped 50% below the long-term norm. More notably, mean air temperatures exceeded the average of direct measurements made since 1851 by more than 6 °C (see ref. 8, for example). According to data derived from the timing of the grape harvest in Burgundy⁹, this heat spell has had no equal since 1370.

Ciais *et al.*¹ have investigated this case by gleaming direct and long-term carbon flux measurements from the CarboEuroflux network and crop-yield data. They also inferred regional-scale carbon fluxes using measurements from the satellite-based Moderate Resolution Imaging Spectroradiometer, and calculations from a regional carbon-cycle model. They report that the 2003 heat spell, combined with the drought, caused a 195 g C m⁻² yr⁻¹ decline in ecosystem photosynthesis and a reduction in ecosystem respiration of 77 g C m⁻² yr⁻¹. And because the decline in respiration was outpaced by the reduction in photosynthesis, Europe experienced a net annual loss of 0.5 × 10¹⁵ g of carbon when the information was integrated across the continent.

The length of the photosynthetic growing season is another temperature-sensitive process that is increasing across Europe¹⁰ and contributes to accumulated photosynthesis. Ciais *et al.*

report that climate conditions preceding the summer of 2003 resulted in an earlier spring-time initiation of photosynthesis compared with 2002, the reference case. This extra carbon assimilation offset the reduction in photosynthesis in the summer of 2003. Whether continued lengthening of the photosynthetic growing season will be correlated with summertime drought and heat remains to be seen. But if it is, it will mitigate the detrimental effects of drought and extreme heat on the annual carbon balance of European ecosystems.

Are these results applicable to similar situations on other continents, and can the consequences be projected into the future? Here we must consider how heat and drought conspire to affect short-term physiological and long-term ecological processes. The discovery that heat and drought stresses reduced both photosynthesis and respiration provides new information for constraining models that couple climate with the carbon cycle. Such models have previously assumed that drought and warming stifle photosynthesis but increase respiration¹¹. In the former situation, perturbations in ecosystem photosynthesis and respiration will increase atmospheric CO₂ and amplify the carbon-cycle feedback on climate warming.

The strength of the observed reduction in the photosynthesis of European forests may not be

P. PAVANI/AFP/GETTY IMAGES

felt by forests in North America under similar conditions of temperature and rainfall. This is because, in North America, temperate forests generally experience higher summertime temperatures than do forests in Europe. Over time, their photosynthetic machinery has acclimated, causing their photosynthetic rates to peak at higher temperatures than for forests in Europe¹². And because the temperature sensitivity of photosynthesis is very plastic, it is reasonable to expect that forests will acclimate if mean temperatures continue to rise gradually across Europe, thereby modulating the negative response to temperature reported by Ciais and colleagues. A repeat of extreme temperatures in the near term, on the other hand, could have detrimental, even lethal, consequences.

One indirect but long-term effect would be a replacement of current forest vegetation with other species. But here again caution is necessary. Many forests in Europe and North America have been regrowing since major harvesting ceased around the end of the nineteenth and beginning of the twentieth centuries. So even if Europe's climate were to remain steady, one may expect certain levels of species transition. Another long-term effect of reduced photosynthesis and growth involves the link between the fall of leaf litter, and its decomposition and provision of nutrients for photosynthesis and plant growth the following year. Ciais and colleagues' study did not last long enough to provide insight into this process, but a reduction in subsequent photosynthesis can be expected.

Finally, the report¹ shows how episodes of heat and drought will affect the ability of European countries to comply with the requirements of the Kyoto Protocol to reduce carbon emissions by limiting fossil-fuel combustion or increasing terrestrial carbon sinks. One potential outcome would be the production of real-time information on carbon cycling, so that fossil-fuel combustion could be adjusted as the weather changes. Achieving such a goal, however, would require an integrated modelling system that predicts weather and carbon cycling in tandem, and the expansion of satellite- and field-measurement systems that would feed such a model.

Dennis Baldocchi is in the Department of Environmental Science, Policy and Management, 137 Mulford Hall, University of California, Berkeley, Berkeley, California 94720, USA.
e-mail: baldocchi@nature.berkeley.edu

1. Ciais, Ph. *et al.* *Nature* **437**, 529–533 (2005).
2. Keeling, C. D. & Whorf, T. P. in *Trends: A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge Natl Lab, Oak Ridge, TN, 2005).
3. Petit, J. *et al.* *Nature* **399**, 429–436 (1999).
4. Meehl, G. A. & Tebaldi, C. *Science* **305**, 994–997 (2004).
5. Harte, J. *et al.* *Ecol. Appl.* **5**, 132–150 (1995).
6. Jones, P. D. & Mann, M. E. *Rev. Geophys.* **42**, doi:10.1029/2003RG000143 (2004).
7. Jones, P. D. & Moberg, A. *J. Clim.* **16**, 206–223 (2004).
8. Stott, P. A., Stone, D. A. & Allen, M. R. *Nature* **432**, 610–613 (2004).
9. Chuine, I. *et al.* *Nature* **432**, 289 (2004).
10. Menzel, A. & Fabian, P. *Nature* **397**, 659 (1999).
11. Cox, P. M. *et al.* *Nature* **408**, 184–187 (2000).
12. Falge, E. *et al.* *Agric. For. Meteorol.* **113**, 53–74 (2002).

STRUCTURAL BIOLOGY

Origins of chemical biodefence

Robert Liddington and Laurie Bankston

The idea that complex biological systems can evolve through a series of simple, random events is not universally accepted. The structure of a vital immune protein shows how such evolution can occur at a molecular level.

Before antibodies evolved, primitive multicellular organisms devised a general defence system against bacterial and viral invaders called 'innate immunity'. The system has survived in vertebrates with its core components little changed during the intervening 700 million years¹. A central element of this defence strategy is an activated thioester — a molecular warhead — that is today used only in this setting, perhaps because it is potentially so destructive. The protein C3, a member of a small family of related proteins carrying this warhead, is a large molecule of the 'complement' system, which identifies foreign agents and targets them for destruction. On page 505 of this issue, Gros and colleagues² present the atomic-resolution crystal structure of C3, as well as that of an inactivated fragment, C3c.

From the structure of C3, it is immediately apparent how this large multi-domain protein evolved from a series of simpler building-blocks into a tightly regulated killing machine (Fig. 1). First, the core of C3 is composed of eight copies of a simple three-dimensional motif, which Gros *et al.* term 'macroglobulin' (MG1–MG8), arranged in tandem. The presence of these repeated domains suggests that the core arose through duplications of a primordial gene that originally encoded a single domain. Random mutations occurring over hundreds of millions of years mean that the component amino-acid sequences of individual domains no longer share any similarity;

nevertheless, their evolutionary origin is preserved in their three-dimensional structure.

Second, the structure shows how domains with specialized functions have been added to the core through a series of gene insertion events. For example, the warhead is contained within the thioester domain (TED), a helical domain that has been crystallized previously as an isolated fragment³. This previous work demonstrates that TED can fold autonomously, and the implication is that it once existed as a separate protein, although there is nothing like it today. TED lies in a loop within a third type of domain, CUB (ref. 4), which in turn fits between the MG7 and MG8 domains of the core. And another domain, called anaphylatoxin (ANA), is inserted in a loop within the MG6 domain.

Why insert the thioester domain in a larger protein? In short: regulation. To be an effective but selective killing machine, C3 must patrol the bloodstream in an inactive form until it encounters 'non-self' (for example, bacteria and viruses). Activation occurs by cleavage of C3 at specific sites (in inserted domains), resulting in the generation of the active fragments, C3b and C3a. By comparing the conformation of full-length C3 with that of the inactivated fragment (C3c) and with the structure of the isolated TED³, Gros and colleagues propose how the molecular warhead of C3 is armed (Fig. 1)². Two steps are involved: first C3 is physically uncovered, then it is chemically primed.

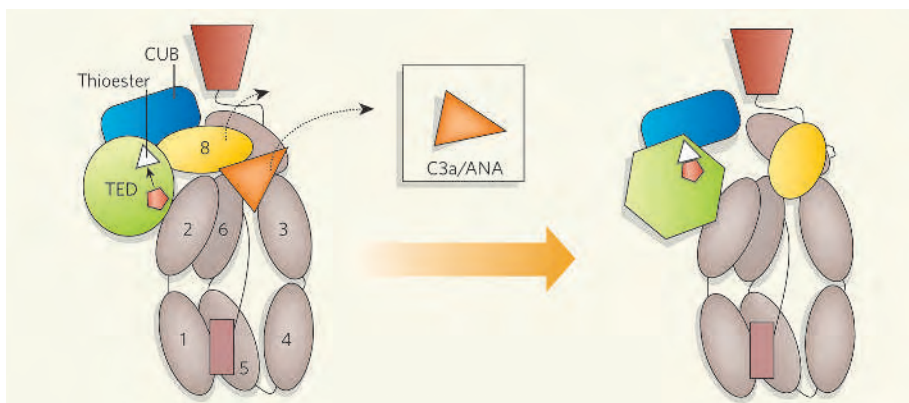


Figure 1 | Arming the warhead. Model of the conformational changes that occur on cleavage and activation of complement component C3 to yield the active fragments C3a/ANA and C3b, based on the structure reported by Gros and colleagues². Proteolytic cleavage of C3 on the pathogen surface releases C3a/ANA from C3b, allowing macroglobulin domain 8 (MG8) to swing away from the thioester domain (TED). This exposes the thioester (white triangle) and enables a switch to an active conformation in which a histidine (red pentagon) reacts covalently with the thioester to form the armed warhead. Further cleavage (not shown) leads to loss of the TED and CUB ('complement C1r/C1s, Uegf, Bmp1') domain, yielding the inactivated fragment, C3c.

In full-length C3, the thioester lies near the surface of the TED domain. However, it is buried by hydrophobic residues of the MG8 domain, effectively excluding water molecules that would otherwise disarm the warhead by hydrolysing the thioester bond. MG8 is held in place by the ANA domain, but following the initial proteolytic cleavage of C3 to form C3b, the ANA domain is released (forming C3a). This in turn frees MG8.

Chemical priming occurs by the reaction between the thioester and the side chain of a histidine amino acid to form a reactive intermediate. However, the histidine is initially too far from the thioester to react, and a conformational change within TED is needed to bring it closer. In the full-length C3, this conformational change is prevented by inter-domain contacts. So the catalytic machinery is inhibited by the multi-domain organization ('quaternary structure') of C3, both by steric blockade and by holding the TED in an inactive conformation. This is an illustration of the concept of allostery first proposed by Monod, Wyman and Changeux⁵.

C3a, once released from C3b, promotes the release of histamine from mast cells and basophils — the start of an inflammatory reaction — and attracts immune cells that gobble up foreigners. Meanwhile, C3b binds to sugars that lie on cell surfaces. In humans, 'self' cells are protected by sialic acid on their surface, which promotes the neutralization of C3b by another complement protein, factor H (ref. 6). But when C3b binds to sugars on the pathogen surfaces, a series of protein–protein reactions and cleavages is set in motion, resulting in the formation of the 'attack' complexes, which bore holes in alien membranes. Binding of active C3b to non-self membranes also triggers their recognition by immune cells that can ingest foreigners.

Various diseases are linked to defects in complement⁷, and, unsurprisingly, pathogens have developed diverse strategies to evade the system⁸, including molecular mimicry of key components⁹. The structures presented by Gros *et al.* not only provide insight into how the complement system evolved, but should also aid the design of drugs to combat disease¹⁰. ■ Robert Liddington and Laurie Bankston are at the Infectious and Inflammatory Disease Center, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA. e-mail: rliddington@burnham.org

1. Sunyer, J. O., Zarkadis, I. K. & Lambris, J. D. *Immunol. Today* **19**, 519–523 (1998).
2. Janssen, B. J. C. *et al.* *Nature* **437**, 505–511 (2005).
3. Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E. & Rini, J. M. *Science* **280**, 1277–1281 (1998).
4. Bork, P. & Beckmann, G. *J. Mol. Biol.* **231**, 539–545 (1993).
5. Monod, J., Wyman, J. & Changeux, J.-P. *J. Mol. Biol.* **6**, 306–329 (1963).
6. Fearon, D. T. *Proc. Natl Acad. Sci. USA* **75**, 1971–1975 (1978).
7. Reis, E. S., Nudelmann, V. & Isaac, L. *Immunogenetics* **55**, 667–673 (2004).
8. Cooper, N. R. *Immunol. Today* **12**, 327–331 (1991).
9. Wallich, R. *et al.* *Infect. Immun.* **73**, 2351–2359 (2005).
10. Hoellers, V. M. & Thurman, J. M. *Mol. Immunol.* **41**, 147–152 (2004).

EARTH SCIENCE

Unleaded high-performance

Tim Elliott

Previous measurements of uranium-series isotopes have implied uncomfortably fast speeds of melt movement through the mantle. Yet the latest results suggest such velocities were serious underestimates.

Most volcanism on Earth occurs unseen, at submarine volcanic ridges that form in response to the seafloor spreading of the oceanic plates. As the plates pull apart at a genteel rate of a few centimetres per year, underlying mantle viscously rises at a similar rate to fill the space. As a result of this upwelling and decompression, the mantle melts, producing the magma that feeds mid-ocean-ridge volcanoes.

On page 534 of this issue¹, Rubin *et al.* provide a dramatic illustration that magma rises to the surface with unexpected haste, in stark contrast to the stately movements of the solid from which it is derived. Melt velocities of up to a few metres per year, about 100 times faster than plate spreading rates, have been rationalized from simple physical models². But Rubin and colleagues' measurements suggest that melts beneath oceanic ridges may move up to three orders of magnitude faster than that — raising questions about our understanding of permeability and fluid flow in the mantle³.

The authors present a geochemical study of very recently erupted mid-ocean-ridge magmas. Investigation of the rates of magma production and transport exploit the uranium-series nuclides, which include isotopes of thorium, radium, radon and lead. These short-lived daughter nuclides occur in the decay chain between the long-lived radioactive parent ²³⁸U (half-life 4.5×10^9 years) and its ultimate, stable daughter, ²⁰⁶Pb. Given time, these intermediate daughter nuclides will establish a steady-state decay chain, in which all nuclides decay at the same rate. This steady state is termed secular equilibrium. Various processes, such as mantle melting, may disturb secular equilibrium, but equilibrium is re-established between any nuclide pair in the uranium-series chain within around five half-lives of the shorter-lived nuclide. Any disequilibrium between a nuclide pair records an event more recent than this. Within the uranium series, nuclide half-lives range from 2.5×10^5 years to 1.6×10^{-4} seconds, providing an ample choice of chronometer.

Compared to many geological processes, the timescales documented by even the longer-lived uranium-series nuclides, such as ²³⁰Th (half-life 76,000 years) and ²²⁶Ra (half-life 1,600 years), are rapid. Yet previous studies of mid-ocean-ridge magmas had already revealed disequilibrium in both ²³⁰Th–²³⁸U and ²²⁶Ra–²³⁰Th nuclide pairs^{4,5}. Rubin *et al.*¹ have upped the ante and analysed ²¹⁰Pb, which has a half life of only 23 years. After

disturbance, the ²¹⁰Pb–²²⁶Ra nuclide pair will return to equilibrium on a timescale of about 100 years, dizzyingly fast for most processes in the Earth's interior.

A major hurdle is to find samples from the seabed that are so young that any ²¹⁰Pb–²²⁶Ra disequilibrium present at eruption has not significantly decayed. It is troublesome to detect, let alone sample, eruptions that occur some 2,500 metres beneath the ocean surface, and so it is a remarkable achievement to obtain lavas to test for initial ²¹⁰Pb–²²⁶Ra disequilibrium. The magmas erupted at mid-ocean ridges also have notably low abundances of uranium and its daughter nuclides, making accurate analysis challenging.

Rubin *et al.*¹ have overcome these problems, and make the striking observation that many of their samples have ²¹⁰Pb–²²⁶Ra deficits — that is, less ²¹⁰Pb than would be expected relative to ²²⁶Ra at the steady state, secular equilibrium. It is both reasonable and conceptually appealing to attribute disequilibrium to the very melting process that produces the magmas. Yet it is also possible that contamination of magma in the crust just before eruption, or degassing of the volatile intermediate ²²²Rn (or even of ²¹⁰Pb itself), produces ²¹⁰Pb deficits. Rubin *et al.*, however, present convincing arguments that these secondary processes do not significantly influence ²¹⁰Pb–²²⁶Ra disequilibrium.

If the ²¹⁰Pb–²²⁶Ra disequilibrium is then a result of melting, rates of melting and melt movement to the sea floor can be inferred. It is first necessary, however, to assess what part of the melting process the disequilibrium is timing. Disequilibrium records the fractionation of parent from daughter nuclide. This occurs during melting because elements have different affinities for melt relative to the melting solid. For example, ²¹⁰Pb enters the melt less readily than ²²⁶Ra, giving rise to a melt with a ²¹⁰Pb deficit. But many of the uranium-series nuclides, including ²²⁶Ra and ²¹⁰Pb, favour the melt over the solid so strongly that the differences in their behaviour are apparent only when very small amounts of melt are present. In simple models of melting, this means that the production of ²¹⁰Pb–²²⁶Ra disequilibrium can occur only when melt is first produced (Fig. 1). The presence of any disequilibrium in erupted lavas then provides a constraint on the time taken for melt to travel from the very bottom of the melting region to the top.

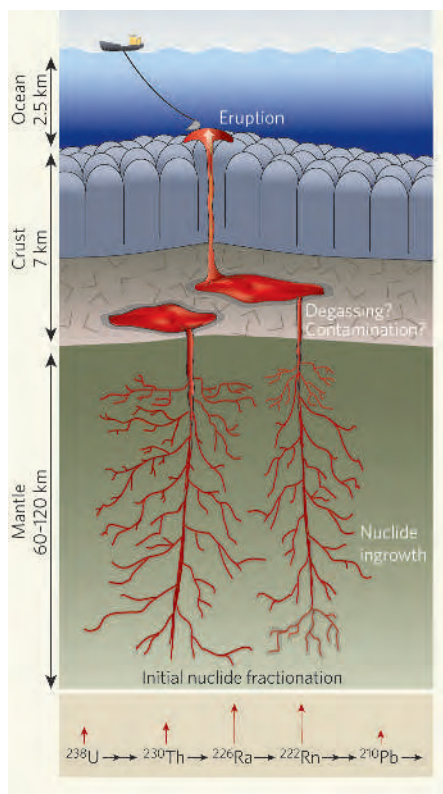


Figure 1 | Melt pathways and possible sites for generation of uranium-series nuclide disequilibrium. Nuclides are fractionated at the onset of melting because of their different affinity for melt relative to the solid. The ^{210}Pb – ^{226}Ra disequilibrium in magmas measured by Rubin *et al.*¹ potentially records this event and so the transit time to eruption. Yet continued equilibration between upwelling melt and solid leads to different velocities for the nuclides through the mantle⁷, generating further disequilibrium. Generation of disequilibrium by such an ‘ingrowth’ process is also only effective where the proportion of the melt is small compared to the solid. That is unlikely to be the case in the main melt conduits, but tributaries to the main channels may contribute ingrown nuclides even high in the melting column. Finally, disequilibrium may be caused by contamination or degassing in the crust, but Rubin *et al.* make a good case against this.

The previous speed limit for this process was clocked in 1988 — also by Rubin, who, together with J. D. Macdougall², used the ^{226}Ra – ^{230}Th pair, which returns to equilibrium in about 8,000 years. The time constraints of this earlier study were thus some two orders of magnitude less stringent than those of the new observations, but at the time they came as a big surprise. In response to the perceived difficulty of moving melt so fast to the surface⁶, melting models were developed that relieved some of the need for speed⁷. However, the less glamorous but quite plausible alternative of crustal contamination has also continually raised its head (Fig. 1).

Importantly, the new study¹ not only requires faster melt transport than before but also provides evidence against some of the increasingly sophisticated scenarios of

contamination⁸. The effects of contamination on the ^{226}Ra – ^{230}Th pair are strikingly different from those on ^{210}Pb – ^{226}Ra . Thus, models constructed to explain previously observed ^{226}Ra – ^{230}Th excesses by contamination seem unlikely to be able to account for the new ^{210}Pb – ^{226}Ra deficits. On the other hand, coupled ^{210}Pb – ^{226}Ra deficits and ^{226}Ra – ^{230}Th excesses are expected for most melting processes. Rubin *et al.* demonstrate that a simple model can reasonably account for their observations.

Clearly, a more comprehensive exploration of the new data using refined models^{9,10} will follow. But now, even more emphatically than before, it seems that you can’t keep a good melt down.

Tim Elliott is in the Department of Earth Sciences, University of Bristol, Wills Memorial Building,

Queen’s Road, Bristol BS8 1RJ, UK.

e-mail: tim.elliott@bristol.ac.uk

1. Rubin, K. H., van der Zander, I., Smith, M. C. & Bergmanis, E. C. *Nature* **437**, 534–538 (2005).
2. Kelemen, P. B., Hirth, G., Shimizu, N., Spiegelman, M. & Dick, H. *Phil. Trans. R. Soc. Lond. A* **355**, 283–318 (1997).
3. Phipps Morgan, J. & Holtzman, B. K. *Geochim. Geophys. Geosyst.* **6**, Q08002; doi:10.1029/2004GC000818 (2005).
4. Condomines, M., Morand, P. & Allègre, C. J. *Earth Planet. Sci. Lett.* **55**, 247–256 (1981).
5. Rubin, K. H. & Macdougall, J. D. *Nature* **335**, 158–161 (1988).
6. Faul, U. *Nature* **410**, 920–923 (2001).
7. Spiegelman, M. & Elliott, T. *Earth Planet. Sci. Lett.* **118**, 1–20 (1993).
8. Saal, A. E. & van Orman, J. A. *Geochim. Geophys. Geosyst.* **5**, Q02008; doi:10.1029/2003GC000620 (2004).
9. Lundstrom, C. *Phys. Earth Planet. Inter.* **121**, 189–204 (2000).
10. Jull, M., Kelemen, P. B. & Sims, K. *Geochim. Cosmochim. Acta* **66**, 4133–4148 (2002).

STRUCTURAL BIOLOGY

Form and function instructions

Jeffery W. Kelly

How much and what kind of information is required to fold a chain of amino acids into a functioning protein? It seems the problem may not be as daunting as once thought — the solution is in the coevolution data.

The linear sequence of amino acids in a protein specifies its final three-dimensional structure and function¹. But what molecular information is necessary and sufficient to specify protein form and function? In papers on pages 512 and 579 of this issue, Ranganathan and colleagues^{2,3} demonstrate that maintaining the conservation pattern in a protein family, along with a surprisingly small subset of coevolving residues, enables the generation of low-homology sequences that fold and function. The studies indicate that the number of crucial interactions in a protein may be smaller than previously thought — a boon for those who want to design novel proteins from scratch to fulfil a specific function.

The authors studied a large family of protein modules, called the WW domains, that mediate protein–protein interactions by binding to sequences that are rich in the amino acid proline⁴. They aligned 120 WW domains from natural proteins, and looked at the distribution of amino acids that occurs at all of the positions along the polypeptide. By comparing each position against the mean distribution from all proteins, they identified those positions that have been conserved throughout evolution and are therefore likely to have some structural or functional significance (Fig. 1). Conservation in these terms means that the amino acid is the same.

They next identified amino-acid positions within the conserved set that seem to have evolved in concert. For example, looking at the

74 sequences with glutamate at position 8 allows comparisons with other positions. Position 16 is a ‘conserved’ position, but it exhibits a mean distribution of residues in the 74 Glu8 sequences, revealing that position 16 is not coupled to Glu8. However, some conserved positions are coupled to Glu8 — that is, they have coevolved. For instance, position 23 has sequence bias in the 74 Glu8 sequences that is distinct from that at the same position in the 120-sequence alignment.

Position 8 is statistically coupled ($P < 0.05$) to just six other sites in the sequence, and a matrix of evolutionary couplings for several positions shows a pattern where a few positions are mutually conserved and most positions interact only weakly, if at all. This result was surprising, and suggested that a computer program might be able to design novel WW domain sequences using conservation and coupling information only, without any information about three-dimensional structure.

To test this idea, Ranganathan and colleagues^{2,3} developed two programs: ‘algorithm 1’ generates artificial WW domain sequences that preserve amino-acid conservation, but eliminates all statistical coupling between sites; and ‘algorithm 2’ generates sequences that retain both statistical conservation and coupling. They produced four libraries of synthetic genes that encode the WW domain. The first library encodes 42 natural WW domains as a control, the second was produced by algorithm 1, the third set was built by algorithm 2,

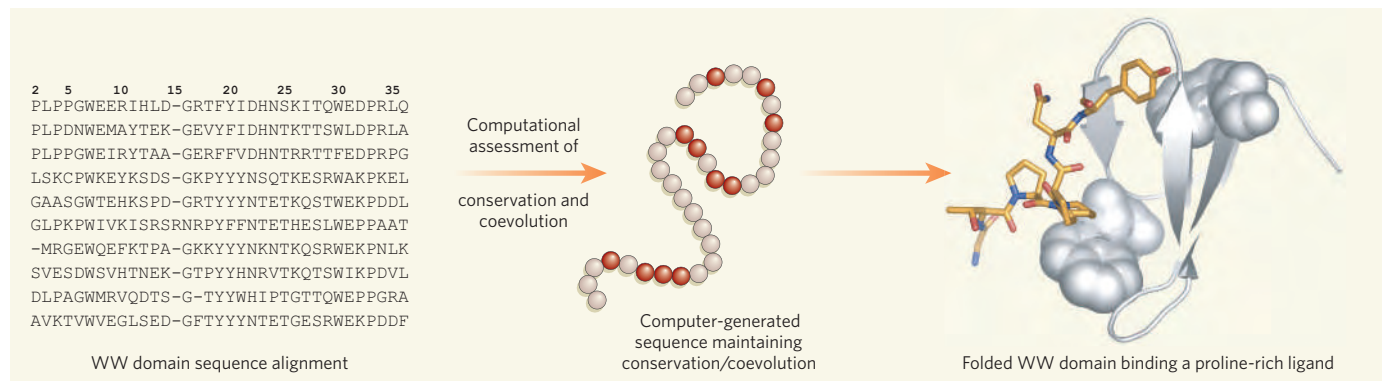


Figure 1 | The grand plan. Ranganathan and colleagues^{2,3} aligned the sequences of WW protein domains and computationally identified those residues showing conservation and coevolution. Computer programs that preserved this information were used to generate libraries of synthetic sequences with modest homology, which were tested experimentally for their ability to fold and function. Conserved and coevolving amino-acid residues are shown as red spheres.

and the fourth set contained 19 random sequences as a negative control.

The authors then used the libraries to express the genes in terms of amino-acid sequence and tested whether they could fold and function like the natural domains. Of the natural sequences, 67% folded in the experimental conditions used. Notably, 25% of the sequences produced from algorithm 2 (conservation + coevolution) could fold, but none of the algorithm-1 sequences (just conservation) could. The sequences generated by algorithms 1 and 2 had analogous extents of identity with natural WW domains, including positions within the hydrophobic core. Moreover, the artificial WW domains function like their natural counterparts: 60% of the algorithm-2 sequences exhibited class-specific binding to proline-rich sequences, with affinities analogous to those seen in many of the original 120 domains.

So it seems that both conservation and evolutionary-coupling information is necessary and sufficient to produce sequences with modest similarity to native sequences that can fold and function. Despite the myriad of local and more-distant interactions revealed by three-dimensional protein structures, these studies imply that there are only a few crucial 'coupled' interactions in a sea of weaker interactions. In other words, there may be a few highly ordered, energetically important molecular interactions that are mixed with many more-fluid, less-important interactions. It is not surprising that these would be difficult to distinguish by simply inspecting averaged structures. It is interesting that many of the conserved and coevolving residues in WW domains are important for fold stability or function^{5,6}. In contrast, the residues that have been shown to control the kinetics of folding are not conserved or coupled, suggesting that folding rate is not optimized by evolution^{5,6}.

Ranganathan and colleagues' algorithm will undoubtedly be tested by the protein-design community in cases for which genetic information is available. The statistical-coupling matrix will allow hypotheses about the evolutionary constraints on a protein to be tested

experimentally. The energetics of folding and function are expected to be manifest as statistical couplings, and it remains to be seen whether these and other features are distinct, overlapping or inseparable features of the total sequence information. It will be interesting to find out whether key conformational changes associated with enzyme and receptor function are preserved in the evolutionary record^{7,8}. Most importantly, this statistical approach, which does not rely on three-dimensional structural information, has the potential to identify crucial residues that perform feats in proteins that we do not currently appreciate. ■

Jeffery W. Kelly is in the Department of Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA.
e-mail: jkelly@scripps.edu

1. Anfinsen, C. B. *Science* **181**, 223–230 (1973).
2. Socolich, M. *et al. Nature* **437**, 512–518 (2005).
3. Russ, W. P., Lowery, D. M., Mishra, P., Yaffe, M. B. & Ranganathan, R. *Nature* **437**, 579–583 (2005).
4. Sudol, M., Chen, H. I., Bougeret, C., Einbond, A. & Bork, P. *FEBS Lett.* **369**, 67–71 (1995).
5. Jager, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. *J. Mol. Biol.* **311**, 373–393 (2001).
6. Deechongkit, S. *et al. Nature* **430**, 101–105 (2004).
7. Shulman, A. I., Larson, C., Mangelsdorf, D. J. & Ranganathan, R. *Cell* **116**, 417–429 (2004).
8. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. *Science* **295**, 1520–1523 (2002).

SYNTHETIC CHEMISTRY

Recipes for excess

John Hartwig

The selective production of a particular mirror-image form of a molecule is immensely important to organic synthesis. But techniques to find the right catalysts have traditionally been protracted and fiddly. Help is at hand.

Chiral molecules are molecules that come in two non-superimposable mirror-image forms, known as enantiomers. Synthesizing one enantiomer of a chiral molecule in preference to the other is difficult but crucial: among other things, single-enantiomer drugs account for some 40% of worldwide drug sales, worth more than US\$100 billion¹. In recognition of this, the 2001 Nobel Prize in Chemistry² was awarded for the development of chiral catalysts for 'enantioselective' synthesis. A notable advance since then has been the recognition that the often neglected, more symmetrical 'achiral' components of a chiral catalyst can be used to improve enantioselectivity further. Recent work^{3,4} that focuses on transformations involving rhodium phosphite and phosphoramidite catalysts is illustrative of this approach.

Living organisms contain countless chiral

biomolecules: proteins, nucleic acids and carbohydrates, not to mention enzyme cofactors, vitamins and other trace substances. On their own, the enantiomers of these molecules — and the enantiomers of small, chiral organic compounds — have the same properties. The distinctive properties of the two enantiomers arise only when they interact with other chiral molecules. This is best explained by the analogy of a handshake. When two people shake hands, they both extend their right hand: a handshake with two right hands fits better than one with a right and a left. Similarly, a 'right-handed' enantiomer of a drug might fit a right-handed 'glove' (binding pocket) of a protein better than the 'left-handed' version of the same drug. The activity of a small-molecule drug (with a molecular mass typically less than 500 daltons) is therefore critically affected by its handedness — its chirality.

One of the most efficient of many approaches to the enantioselective preparation of a small molecule is to use a catalyst that is itself chiral, usually comprising a chiral group, or ligand, bound to a central metal atom. But chemists' ability to predict the structure of a ligand that will help to synthesize a product with high enantioselectivity is limited. This is because the difference in the size of the energy barriers that must be overcome to form each enantiomer — even in substantially different amounts — is typically small relative to the size of the barriers. In practice, ligands are often tested by trial and error. Unfortunately, the selective synthesis of one enantiomer of a ligand is itself challenging, and the synthesis of a series of such ligands is usually the slowest step in the development of enantioselective catalysts. Progress therefore depends on the improved ability to control small differences in energy, on developing better methods to produce single enantiomers of chiral ligands, or on developing approaches that bypass, at least partially, the need to produce sophisticated chiral ligands.

Consequently, the focus of some efforts to design enantioselective catalysts has in recent years shifted away from chiral ligands and towards achiral ligands. These ligands are either bound to the central metal atom of the catalyst, like a chiral ligand, or can adopt a chiral arrangement at the appropriate stage in the catalytic cycle (for a review of these developments, see refs 5, 6). Such approaches in effect subdivide the catalyst's structure, and allow pieces of the catalyst to be varied in a modular fashion, instead of varying portions of whole molecules that must be made one at a time.

Several years ago, Walsh and co-workers demonstrated that zinc complexes containing both an achiral and a chiral ligand react with higher enantioselectivity, and at higher rates, than do complexes containing only the chiral ligand^{7–9}. Varying the size and shape of the achiral catalyst components — much as one might vary the shape and size of the (achiral) fingers on a (chiral) hand to accomplish an intricate physical task — leads to large swings in the excess of one enantiomer product over the other. It can even cause the enantiomer formed using the same combination of zinc and chiral ligand to switch from left-handed to right-handed, or vice versa. Even higher enantioselectivity has been demonstrated by reactions conducted with zinc binolate complexes in combination with a ligand that is achiral when free from the zinc, but chiral when bound⁸. Such combinations of chiral and achiral ligands should allow the rapid, parallel synthesis of families of potentially enantioselective catalysts.

These techniques have now been used^{10–12} to develop catalysts for hydrogenation, the most common class of reaction for preparing single-enantiomer drugs. Rhodium complexes bound by two separate chiral ligands catalyse hydrogenation reactions with enantioselectivities

rivalling those achieved by catalysts comprising two ligands tethered together. Significantly for practical applications, derivatives of these new catalysts can be prepared simply by mixing two different chiral ligands, or a chiral and an achiral ligand, with the rhodium catalyst precursor.

In a similar vein to the work^{6–9} on zinc complexes, Reetz and Li³, writing in *Angewandte Chemie International Edition*, investigate the formation of catalysts that combine a variety of phosphorus-based chiral and achiral ligands. In the same journal, Feringa and colleagues (Hoen *et al.*)⁴ focus on a combination of chiral phosphoramidite and achiral tertiary phosphine ligands. Both groups find that, as with zinc complexes, enantioselectivity is in some cases higher in the presence of achiral ligands than in their absence.

Reetz and Li³ take this design further by combining a single enantiomer of a chiral phosphite ligand with a second phosphite that is an equal, 'racemic' mixture of two enantiomeric conformations when free in solution. They find that the single enantiomer of one chiral phosphite can dictate the chiral configuration of the second phosphite; the combination of these two ligands can then preferentially form one enantiomer of the reaction product. The authors use this process to generate chiral amines, which are substructures of a host of pharmaceutical candidates, from hydrogenation of an enamide, with high enantioselectivity.

Although the focus on varying both chiral and achiral ligands is new, it borrows from the first successes in enantioselective catalysis. There, the enantioselectivity of a catalyst was generally controlled by varying achiral component groups, such as flat benzene-like aromatic rings, in a chiral ligand. The array of aromatic rings created a chiral object, much as flat propellers — themselves symmetrical and superimposable (and so achiral) — combine

to form a symmetrical, but non-superimposable (chiral) object such as a ceiling fan. The size and the shape of these aromatic propeller blades affect the chiral structure of the metal active site and so the enantioselectivity of the catalyst.

The developments outlined here^{3,4,7–9}, and others¹³, which combine a chiral ligand with a second achiral ligand, developed out of attempts to avoid the laborious and time-consuming nature of the sequential modification of these achiral groups. This simplified approach to preparing catalysts for enantioselective chemistry significantly increases the number of catalysts that can be tested. The chances of discovering highly enantioselective catalysts for useful chemical reactions — including those for developing new drugs — are thus vastly increased. ■

John Hartwig is in the Department of Chemistry, Yale University, PO Box 208107, New Haven, Connecticut 06520-8107, USA.
e-mail: john.hartwig@yale.edu

1. Stinson, S. C. *Chem. Eng. News* **79**, 79–97 (2001).
2. <http://nobelprize.org/chemistry/laureates/2001/index.html>
3. Reetz, M. T. & Li, X. G. *Angew. Chem. Int. Edn* **44**, 2959–2962 (2005).
4. Hoen, R. *et al.* *Angew. Chem. Int. Edn* **44**, 4209–4212 (2005).
5. Vogl, E. M., Gröger, H. & Shibasaki, M. *Angew. Chem. Int. Edn* **38**, 1570–1577 (1999).
6. Walsh, P. J., Lurain, A. E. & Balsells, J. *Chem. Rev.* **103**, 3297–3344 (2003).
7. Walsh, P. J., Balsells, J., Betancort, J. M. & Gama, G. J. *J. Am. Chem. Soc.* **122**, 1802–1803 (2000).
8. Costa, A. M., Jimeno, C., Gavenonis, J., Carroll, P. J. & Walsh, P. J. *J. Am. Chem. Soc.* **124**, 6929–6941 (2002).
9. Lurain, A. E., Carroll, P. J. & Walsh, P. J. *J. Org. Chem.* **70**, 1262–1268 (2005).
10. van den Berg, M. *et al.* *J. Am. Chem. Soc.* **122**, 11539–11540 (2000).
11. Reetz, M. T. & Mehler, G. *Angew. Chem. Int. Edn* **39**, 3889–3890 (2000).
12. Reetz, M. T., Sell, T., Meiswinkel, A. & Mehler, G. *Angew. Chem. Int. Edn* **42**, 790–793 (2003).
13. Miller, J. A., Gross, B. A., Zhuravel, M. A., Jin, W. & Nguyen, S. T. *Angew. Chem. Int. Edn* **44**, 3885–3889 (2005).

TECHNIQUES

Imaging at a distance

Siegfried Stapf

Magnetic resonance imaging is often limited by the need to encode information and acquire the resonance signals in less-than-ideal locations. Performing these two steps at different places provides a solution.

Nuclear magnetic resonance (NMR) is an outstandingly versatile technique, used across the sciences. The basic principles — manipulating the nuclear spins of atoms in a sample by means of radio waves and magnetic fields, and recording the resonance signal obtained — can be applied to provide information in such disparate fields as chemistry and medical diagnostics. Yet the method has its limitations, leaving scope for further innovation. One such

advance, reported in *Physical Review Letters*¹, stems from the work of Pines and colleagues.

The problem they have tackled is that of obtaining sharp images of fluid flow inside a rock. Studies of this kind of sample are limited by the many interfaces and, in some cases, the metal content of natural rocks, which bend the applied magnetic field and blur the magnetic resonance image. NMR is used in oilfield prospecting and provides valuable

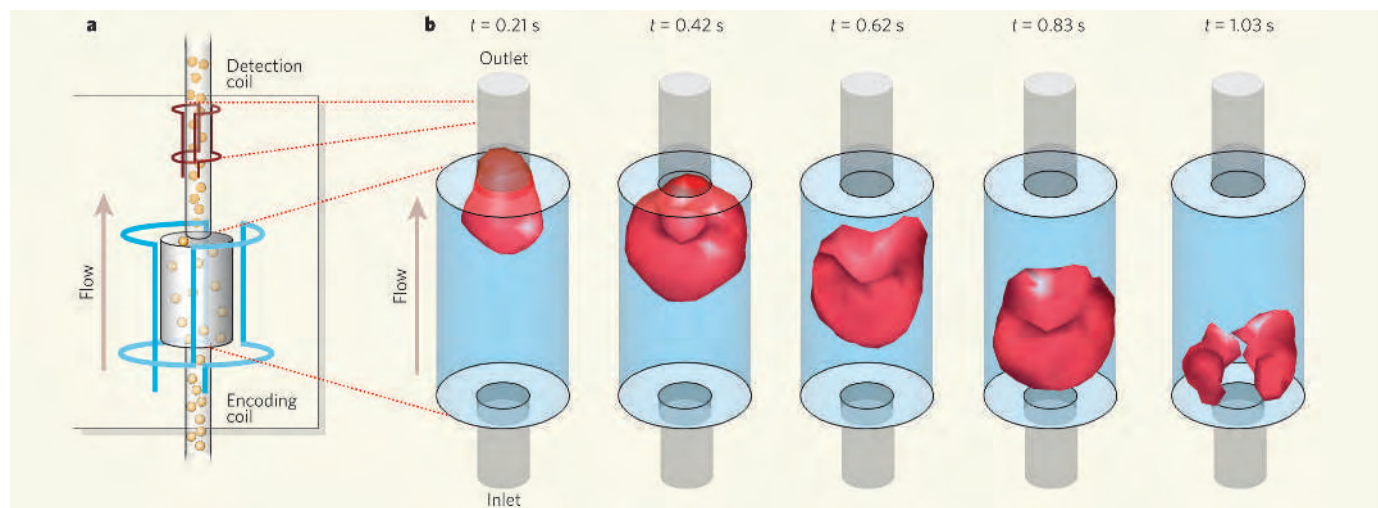


Figure 1 | Remote-detection NMR. **a**, In this approach, the first step (excitation of spins in the sample by the encoding coil) is separated in space from the second step (detection of the resonance signal in the detector coil). **b**, Results from an experiment on a rock sample¹. Following spin manipulation of the gas mixture inside the entire 38-mm-long sample, spin-density images (red) were acquired at different time delays at the remote detector downstream from the sample's end. Short delays indicate molecules close to the sample outlet; increasing delays bring those from near the inlet into view. In this way, the pattern of fluid residence times, and heterogeneities owing to fractures or isolated pores, can be deduced. (Figure modified from ref. 1.)

information about the oil and water content of rocks. But taking a closer look at the fluid distribution remains a difficult task, even under laboratory conditions.

Similar obstacles dog the application of NMR to samples that are very large, very small or highly diluted, and researchers have long tried to adapt their hardware to cope with the field distortions induced by such samples. Pines and colleagues² have previously questioned the view that the whole NMR procedure needs to take place in one location. They aimed at obtaining a sharper view of the object under study by the counterintuitive approach of moving away from it. The two steps of manipulating the spins and acquiring the resonance signal are usually combined in the NMR apparatus, but in the authors' 'remote-detection' approach they are separated not only in time, but also in space. Excitation of the spins occurs at one place and the signal is collected at a different location, with magnetic fields and detectors independently optimized in the two steps to maximize signal quality (Fig. 1a).

For their latest work¹, Pines and colleagues joined forces with researchers at Schlumberger-Doll, a company that provides technical advice and NMR hardware for the oil industry. The sample they used was a small cylinder of sandstone rock, 20 mm in diameter and 38 mm long, and the test fluid was a gas mixture that included the isotope xenon-129. In the first stage of the procedure, the nuclear spins in the gas were manipulated inside the encoding coil by a combination of radio-frequency pulses and varying magnetic-field gradients. As a result, the gas had a 'memory' of its position at the time of spin encoding. In the second step, fluid flow carried the gas through the sample and then past a small, highly sensitive detector coil, where the stored NMR information was collected.

Notably, the authors modified their original concept by exploiting the information contained in the travelling time to the detector. Each image acquired with a particular delay provides a snapshot of the fluid distribution at the corresponding instant (Fig. 1b). The molecules closest to the sample outlet are the first to pass the detector coil; by contrast, those trapped in dead-end rock pores remain undetected, and so heterogeneities in the pore space can be identified.

The ultimate goal of this particular line of research is to be able to perform such experiments in the field. By lowering an entire NMR spectrometer inside a small borehole, the residence-time distribution of a fluid inside the rock might identify pockets of oil that would be too costly to recover. So the technique could potentially provide more complete information than can be obtained by even a full three-dimensional rendering of the pore space. Problems to be overcome before this goal can be achieved include the small space available for the probe and the limited timeframe for executing the measurement while drilling. The currently used inside-out configuration for NMR — where the sample is outside the instrument — can be modified by Pines and colleagues' two-coil design so that the otherwise weak signal is 'concentrated' into a small and efficient receiver coil.

Although the experiment used gas as the test fluid, liquids could be used. However, xenon gas has the advantage of ensuring that the spin magnetization has a long enough lifetime to survive the transport process. Moreover, xenon-129 is extremely sensitive to its molecular environment³ and its spin polarization can be boosted dramatically by optical methods⁴, making it a popular tracer in NMR studies. The distributions and transport properties of gases, however, are notoriously difficult to detect by conventional approaches

because of their low density, and here the two-coil design pays dividends because the gas can be collected outside the sample.

The concept of remote detection is being explored further, opening up other possibilities. For instance, combining NMR with advanced spin-detection technologies, such as using superconducting quantum interference devices, or SQUIDs^{5,6}, becomes feasible. Such a prospect had previously seemed unpromising because of the restrictions of conventional NMR conditions. More achievable applications might exploit the advantages of small magnetic fields for the encoding step²; even Earth's weak magnetic field can suffice⁷, because magnetic couplings between nuclei are independent of field strength and become the dominant source of information about molecular structure in small external fields.

Finally, NMR remote detection has promising applications in chemical engineering. Here, the same issue of the imaging of flow through tiny channels occurs in the microreactors used in fine chemical synthesis. Recently, Pines and colleagues have monitored fluid residence times inside a microreactor, and find that the remote-detection approach might well be used to optimize transport and reaction efficiency in such microstructured devices⁸.

Siegfried Stapf is at the Institute of Technical Chemistry and Macromolecular Chemistry, RWTH Aachen, 52074 Aachen, Germany. e-mail: sstapf@mc.rwth-aachen.de

1. Granwehr, J. et al. *Phys. Rev. Lett.* **95**, 075503 (2005).
2. Moule, A. J. et al. *Proc. Natl Acad. Sci. USA* **100**, 9122–9127 (2003).
3. Miller, K. W. et al. *Proc. Natl Acad. Sci. USA* **78**, 4946–4949 (1981).
4. Happer, W. et al. *Phys. Rev. A* **29**, 3092–3110 (1984).
5. Seeley, J. A., Han, S. & Pines, A. *J. Magn. Reson.* **167**, 282–290 (2004).
6. Greenberg, Y. S., *Rev. Mod. Phys.* **70**, 175–222 (1998).
7. Appelt, S., Häsing, F. W., Kuhn, H., Perlo, J. & Blümich, B. *Phys. Rev. Lett.* **94**, 197602 (2005).
8. Wensink, H. et al. *Lab on a Chip* **5**, 280–284 (2005).

NEWS & VIEWS FEATURE

PHARMACEUTICALS

A new grammar for drug discovery

Mark C. Fishman and Jeffery A. Porter

To realize the potential of the genome for identifying candidate drugs we must move beyond individual genes and proteins. The signalling pathways in cells provide the right level for such analyses.

Drug discovery is a tough business scientifically, the traditional steps taking nearly a decade. In this article we are concerned only with the very beginning of the process, the identification of a target for a new drug. It is this step, the choice of a gene product of clinical relevance, that is the greatest impediment to expanding the pharmaceutical arsenal. In the United States, only 20–30 new chemical entities are approved as drugs each year, and the picture is much the same in Europe. Of these, only a quarter act on targets not already hit by an existing drug. Why has the deciphering of the 25,000 or so genes in the human genome not swelled the ranks of new targets?

The major reason, of course, is that targets are of value for drug discovery only if they can be convincingly related to disease. Such validation really takes not the one year shown on standard pipeline charts (Fig. 1), but decades. For example, the discovery of statins as agents that lower cholesterol levels in the bloodstream rested on investigations that began 40 years earlier with a study showing the relationship between cholesterol level and vascular disease¹. The recent discovery of Gleevec, which has revolutionized the therapy of chronic myeloid leukaemia, certain gastrointestinal tumours and other cancers, was based on work that began in the 1950s (ref. 2). Thus, it has been argued that well-validated targets, the low-hanging fruit as it were, are simply exhausted^{3,4}.

Validation is generally not a one-step process, but is rather an edifice built from studies in epidemiology and disease physiology, and from the results of research with animal models. The one exception is mendelian disease. In these disorders, such as cystic fibrosis and sickle-cell anaemia, the inheritance of a mutation in a single gene can be incontrovertibly linked to a physical characteristic, or phenotype — in this case the disease. Of the 6,000 or so illnesses with a mendelian pattern of inheritance, the gene responsible has been identified in approximately 1,200 (refs 5, 6). Historically, pharmaceutical companies have not concentrated on these diseases, in some cases because the affected protein is not tractable to pharmaceutical approaches, in others, perhaps, because the number of people affected is small. But the

powerful role of a single gene in mendelian disease can provide insight into complex diseases where the same gene accounts for part of the phenotype. Statin therapy, for example, was initially directed to patients with a genetic predisposition to excessive levels of blood cholesterol. But after the drug's efficacy and safety had been tested, the treatment was extended to a wider population of patients who had the same condition but due to many causes^{1,7,8}.

In seeking new targets, we first need to think about how we describe disease. Medical textbooks are organized by organ system, and diseases are classified by their pathology or physiology. If we hope to come to grips with the heterogeneity of common disease and its consequences for the choice of targets for drug discovery, clinical manifestations of disease must be causally related to a molecular definition. A simple concatenation of all the molecular changes in a disease, for example a compilation of profiles of which genes are being transcribed, and when and where, would be chaotic. A further step of integration and interpretation is needed. Ideally, this would provide a 'grammar' that would apply right through from the discovery of a drug target, to its testing in animal models and finally to the treatment of patients. In this context, a grammar means a set of rules, not for organizing words into sentences but for translating gene products into medicines. Just as the

grammar we use in our speech is embedded in language, a grammar of drug discovery should be embedded in biological systems.

Molecular pathways

Intracellular molecular signalling pathways provide such a grammar for the genome. These pathways are triggered by extracellular molecules that bind to receptors in the cell membrane, thereby switching on relay systems inside the cell. The upshot is gene activation, or inactivation, that affects a cell's behaviour — its ability to grow or differentiate, for example, or to undergo division or self-destruction. The number of distinct signalling pathways recognized in human cells depends on the definition. If classified by the type of cell-surface receptor involved, it could be as low as 16 (ref. 9). Using definitions based on cell type or gene-family member, and variants thereof, it could be as many as 200 (ref. 10). In any case, the number of signalling pathways is far fewer than the 25,000 or so human genes.

The elements of four now-canonical molecular pathways are shown in Figure 2 (overleaf): these lead from the receptors for proteins of the Wnt, Hedgehog (Hh), transforming growth factor- β (TGF- β) and insulin/insulin-like growth factor (IGF) families. Only the core components are depicted here in a much simplified manner. The important point is that these pathways are conserved throughout

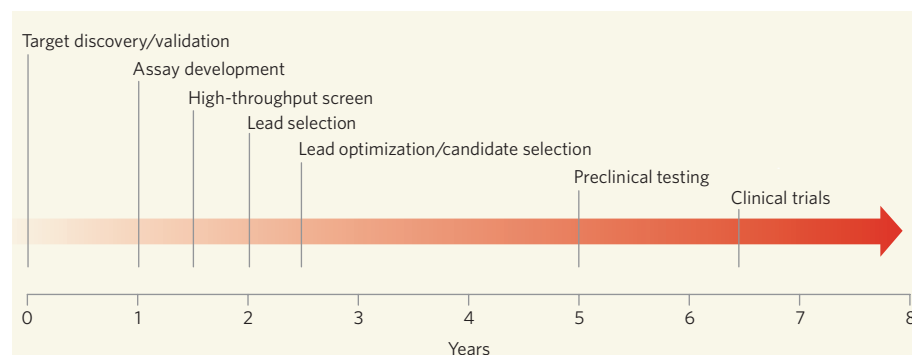


Figure 1 | Drug-discovery timeline. The initial phase, where promising targets are first discovered and linked to a disease, is shown here as taking a year, but it can require several years of work in both university and industrial labs. Subsequent stages involve screening for and identifying compounds that safely alter the activity of a particular target, before their efficacy is tested in clinical trials.

most of the animal kingdom, in invertebrates and vertebrates, and so they evidently control some of the basic cellular functions of life.

The power of molecular pathways to provide a grammar is evident from studies of embryonic development. As first demonstrated in the fruit fly *Drosophila*, the language of development is best phrased by the grammar of molecular pathways¹¹. At a molecular level, the characteristic units (sentences, if you will) of fly development are written in the form of pathways that determine, for example, the first anterior–posterior patterning decision, or the subsequent division of the body plan into segments or the generation of wings, eyes and other tissues. In zebrafish and mouse, similar unitary modules, often using the same molecular pathways, also define elements of organ systems and function in embryonic development¹².

Fundamental cellular processes that control growth and cell death in adults are similarly defined by canonical molecular pathways^{13,14}. Such pathways can be represented graphically¹⁰ and measured quantitatively¹⁵. The order and interactions of pathway components can be defined by genetics and by protein chemistry. All in all, they can provide a grammar applicable to drug discovery and medicine, from target validation through to the clinic.

The link between disease and signalling pathway is best validated for genetic disorders. Table 1 lists several diseases caused by mutations in the genes for components of the Wnt, Hh, TGF- β and insulin/IGF pathways. Perturbation of the essential processes driven by these pathways is the cause of many diseases, including diseases with complex causes, such as diabetes and heart disease.

Perhaps the best example known so far is the perturbation of a single pathway that regulates cell growth — the insulin/IGF–AKT pathway (Fig. 2) — in a host of apparently unrelated diseases, and how targeting of a key node in this pathway can have beneficial effects in all those diseases¹³. The pathway is activated in many cancers. For example, an increase in levels of phosphatidylinositol 3-kinase (PI3K) or mutations in its regulatory subunit cause cancers of the ovary and stomach, and activation of AKT or a decrease in PTEN function cause tumours of the prostate gland.

Other growth phenomena also depend on the insulin/IGF pathway. Examples are the proliferation of smooth muscle that leads to a narrowing of arteries (restenosis) following treatment to unblock them, and the increase in the numbers of certain immune cells — T cells — during transplant rejection. A key downstream node in this pathway in all these disorders is the protein mTOR, which regulates the translation of messenger RNA into protein. In clinical trials, inhibition of mTOR with derivatives of the macrolide rapamycin, an immunosuppressive drug, is now proving effective against cancer and restenosis as well as in transplant rejection.

The key prediction of such logic is that the

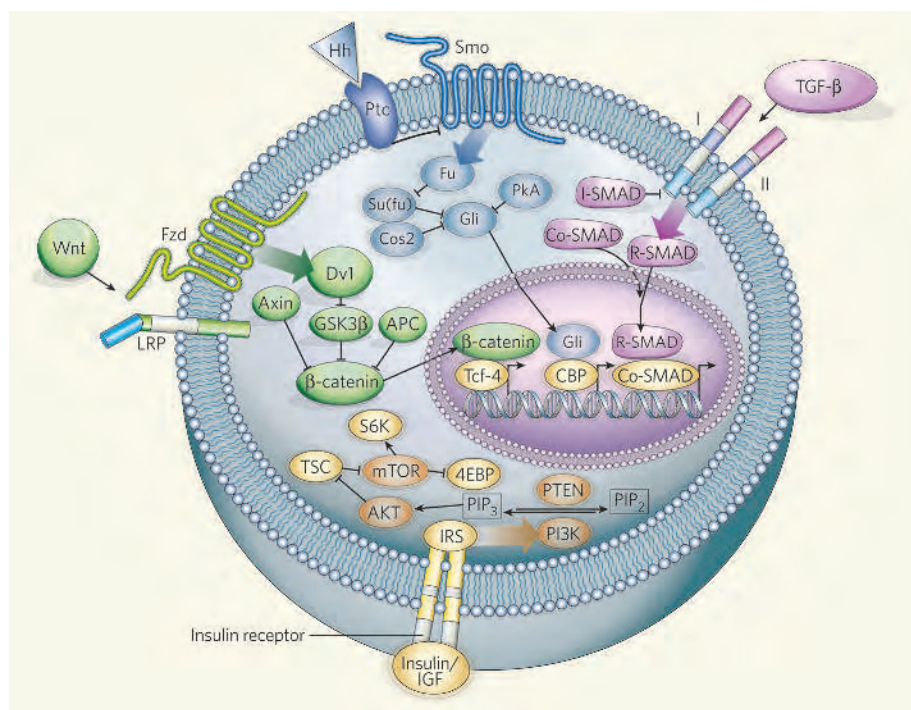


Figure 2 | Four signalling pathways linked to disease. The diagram depicts the main components of the Wnt, Hedgehog (Hh), transforming growth factor- β (TGF- β) and insulin/insulin-like growth factor (IGF) pathways within a single cell. In each case, triggering a receptor on the cell surface activates molecular relay systems, which, on transmission of the signal into the nucleus, result in genes being switched on or off. The arrows indicate activation and the T-bars show inhibition. Misregulation of several of these components is directly associated with various inherited and sporadic illnesses (see Table 1). For simplicity, details of the component nomenclature and known interactions are not covered here (see ref. 10).

drug target can be a ‘sensitive node’, not necessarily the protein that is specifically perturbed — which may be hard to define or not readily ‘druggable’ using available treatments. For example, predisposition to colon cancer is due in some cases to inherited mutations in APC, a component of the Wnt pathway (Fig. 2), a protein that is technically difficult to target. But compounds that block downstream interactions in this pathway, those of β -catenin with its co-activators, have been discovered, and are able to suppress growth of tumours arising because of mutations in APC^{16,17}.

How can we discover the druggable nodes in other pathways? One approach is to select points of intervention, based on the current understanding of the pathways and their druggable components, and begin building cell-free assays to screen for small molecules that will alter the activity of these targets. A less biased approach is to combine genetics with chemical screens using potential drugs. The premise of such an approach is that the effect of a drug resembles that of the genetic perturbation of its target^{18–20}. For example, screens in yeast have shown that the growth-inhibitory effect of a compound often matches that of a mutation in the gene encoding its target^{21,22}. In the chemical screen, libraries of drug-like compounds are applied to cells, or even whole organisms, to alter pathway activity, and their effects are then compared with those of known mutations. The presumptive target can be confirmed

biochemically. Thus, in principle, a compendium of drugs could be established that interrupt or stimulate every pathway of interest.

Next steps

‘In principle’ is the crucial phrase here, for our current understanding of molecular pathways is insufficient as a platform for effective pharmaceutical discovery. Several biotechnology companies have focused on the known elements of a few key pathways to target them with new medicines. But for the genome to be translated into medicines with any reliability and regularity, far more work needs to be done. Defining the role of pathways in complex diseases will undoubtedly take many years. A first step is to define the full complexity of these signalling networks at a molecular level — their ‘systems biology’ — including activity specific to a particular cell type, dynamic feedback mechanisms, extensive inter-pathway connectivity, the kinetics of signalling, and, of course, their state of activation in disease. The following issues will have to be addressed.

● Pathway outputs have different effects in different contexts. For example, the pathway activated by another extracellular signal, fibroblast growth factor (FGF), controls cell division in many tissues but in others it regulates cell differentiation, shape or survival²³. Sonic hedgehog, one of the related Hh signalling

molecules, defines cell fates in the embryonic neural tube, but promotes cell division in the embryonic cerebellum²⁴. The outcome may even vary for a single tissue. For example, when it acts in breast epithelium, TGF- β normally helps to prevent the malignant transformation of the cells, in that it hinders cell growth and multiplication and promotes cell suicide. But if the epithelial cells are transformed as a result of a mutation upstream of SMAD, for example, TGF- β enhances blood-vessel formation and tumour-cell invasiveness, which encourage tumour growth and dispersal²⁵. Understanding the druggable elements or nodes in pathways that can trigger such altered biological states should define new targets for treating many different diseases.

● Pathways are dynamic, with feedback mechanisms. For example, FGF stimulation of airway formation in *Drosophila* is terminated by the intracellular activation of the protein Sprouty, which restricts the process to a specific location^{23,26}. Similarly, Wnt signalling causes the up-regulation of inhibitors of the Wnt receptor Fzd, which then inhibit the pathway (Fig. 2)²⁷. In organ systems, changes in feedback, or in its phasing with regard to the stimulus, can have adverse effects. For example, abnormal breathing patterns result when feedback to the chemoreceptor concerned is out of phase with its output. It is likely that feedback abnormalities have adverse consequences in a cellular context as well. A major goal is to define the molecular components that are responsible for maintaining physiological balance, as these proteins are likely to be key regulators of disease.

● Pathways intersect. Work in classical genetics has defined the phenotypic consequences when different sites on the same pathway are perturbed¹¹. Pathways also engage in extensive cross-talk^{28,29} with each other, however, and these interactions change over time, for example after stimulation. But despite the complexity of these interactions, they are finite: they can be mapped and the key nodes in the intersections identified. As we come to understand

this cross-talk and its role in disease, a logic for combination therapy should emerge in which two or more pathways can be targeted simultaneously.

● Pathway kinetics vary between cells. The quantitative flow of signals through pathways depends on the levels of the specific components of a pathway. For example, quantitative modelling shows that the level of axin (Fig. 2) is what determines the amplitude and duration of signals through the Wnt pathway¹⁵. To create cell-specific modulators of pathways we will probably need to define these limiting steps and target them.

Clinical relevance

One advantage of a pathway-based clinical taxonomy is that it affords a mechanistic foundation for disease description. Historically, diseases have been categorized by organ pathology. Today we are moving towards pan-genomic assays of gene expression, protein levels and the ways in which proteins become modified after they have been produced from mRNA. The results of these assays become easier to interpret when several elements of a pathway are affected. Subtle changes in an individual protein may be below the threshold of detection, but may be amplified by sophisticated computational methods that incorporate several pathway components³⁰.

Another advantage of focusing on pathways is the ability to choose the most appropriate set of patients for initial tests of possible treatments. For many pathways, people with specific genetic defects have been described (Table 1), but, because of their rarity, the discovery of a treatment has often been relegated to the 'orphan drug' category. The development of therapies for such patients would not only serve a medical need, but would often be readily extrapolated to a wider population. For example, Gorlin's syndrome, which results from mutations in the protein Patched-1 (Ptc in Fig. 2; a negative regulator of Hh signalling), causes a brain cancer called medulloblastoma and a common form of skin cancer. The syndrome

itself only affects between 1 in 50,000 and 1 in 150,000 people³¹. But many other tumours, including non-small-cell lung cancer, gut-related tumours and prostate cancer, depend on Hh signalling, even though they do not seem to have mutations in pathway components³².

The rationale for extrapolating from genetic to sporadic illness is that nature is conservative: this is a safe bet. And there would be immense immediate benefit in tackling the rare diseases in themselves. Patients with uncommon disorders are often neglected, their only hope being that medicines designed and marketed to treat common disorders might coincidentally be able to treat theirs. In contrast, in the logic of molecular pathways, such patients would be viewed as key intermediaries in the drug-discovery process, providing proof of concept. Along the undoubtedly long route to making molecular pathways a useful grammar for medicine, essential steps will involve successfully treating these well-defined but rare diseases. This approach could not only bring a new order to the genome, but could also have a salutary 'side effect' — a refocusing of drug-discovery research on neglected diseases. ■

Mark C. Fishman and Jeffery A. Porter are at the Novartis Institutes for BioMedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

e-mail: mark.fishman@novartis.com

1. Tobert, J. A. *Nature Rev. Drug Discov.* **2**, 517–526 (2003).
2. Nowell, P. C. & Hungerford, D. A. *Science* **132**, 1487–1501 (1960).
3. Drews, J. *Science* **287**, 1960–1964 (2000).
4. Brown, D. & Superti-Furga, G. *Drug Discov. Today* **8**, 1067–1077 (2003).
5. Hamosh, A. et al. *Nucleic Acids Res.* **30**, 52–55 (2002).
6. Schuener, M. T., Yoon, P. W. & Khuury, M. J. *Am. J. Med. Genet.* **125C**, 50–65 (2004).
7. Mabuchi, H. et al. *N. Engl. J. Med.* **305**, 478–482 (1981).
8. Mabuchi, H. et al. *N. Engl. J. Med.* **308**, 609–613 (1983).
9. Gerhart, J. & Kirschner, M. W. *Cells, Embryos, and Evolution* (Blackwell, Malden, MA, 1997).
10. www.biocarta.com/genes/allPathways.asp
11. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795–801 (1980).
12. Harris, T. E. & Lawrence, J. C. Jr *Science STKE* 2003, re15 (2003).
13. Metzstein, M. M., Stanfield, G. M. & Horvitz, H. R. *Trends Genet.* **10**, 410–416 (1998).
14. Fishman, M. C. & Olson, E. N. *Cell* **91**, 153–156 (1997).
15. Lee, E., Salic, A., Kruger, R., Heinrich, R. & Kirschner, M. W. *PLoS Biol.* **1**, 116–132 (2003).
16. Emami, K. H. et al. *Proc. Natl Acad. Sci. USA* **101**, 12682–12687 (2004).
17. Lepourcelet, M. et al. *Cancer Cell* **5**, 91–102 (2004).
18. Stockwell, B. R. *Nature Rev. Genet.* **1**, 116–125 (2000).
19. Giaever, G. et al. *Nature Genet.* **21**, 278–283 (1999).
20. Hughes, T. R. et al. *Cell* **102**, 109–126 (2000).
21. Lum, P. Y. et al. *Cell* **116**, 121–137 (2004).
22. Parsons, A. B. et al. *Nature Biotechnol.* **22**, 62–69 (2004).
23. Tsang, M. & Dawid, I. B. *Science STKE* 2004, pe17 (2004).
24. Ruiz i Altaba, A., Nguyen, V. & Palma, V. *Curr. Opin. Genet. Dev.* **13**, 513–521 (2003).
25. Siegel, P. M. & Massagué, J. *Nature Rev. Cancer* **3**, 807–821 (2003).
26. Hacohen, N., Kramer, S., Sutherland, D., Hiromi, Y. & Krasnow, M. A. *Cell* **92**, 253–263 (1998).
27. Logan, C. Y. & Nusse, R. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).
28. Bouwmeester, T. et al. *Nature Cell Biol.* **6**, 97–105 (2004).
29. Giot, L. et al. *Science* **302**, 1727–1736 (2003).
30. Mootha, V. K. et al. *Nature Genet.* **34**, 267–273 (2003).
31. www.emedicine.com/ped/topic890.htm
32. Berman, D. M. et al. *Nature* **425**, 846–851 (2003).

Pathway	Associated diseases	Affected gene
Hedgehog	Gorlin's syndrome	Ptc1
	Basal cell carcinoma	Ptc1
	Medulloblastoma	Ptc1
	Glioblastoma	Gli1
Wnt	Vitreoretinopathy	Fzd4
	Norrie's disease	NDP
	Colon	APC
	Osteoporosis-pseudoglioma/osteopetrosis	LRP5
TGF- β	Fibrodysplasia ossificans	BMP4
	Haemorrhagic telangiectasia	Alk1
	Pulmonary hypertension	BMPR2
	Juvenile polyposis syndrome	SMAD4
	Pancreatic cancer	SMAD4
Insulin/IGF	Cowden's disease	PTEN
	Tuberous sclerosis	TSC1
	Multiple cancers	PTEN, PI3K, AKT

Table 1 | Fundamental pathways and a partial list of disease links.

BRIEF COMMUNICATIONS

'Devil's gardens' bedevilled by ants

An ant species uses herbicidal weaponry to secure its own niche in the Amazonian rainforest.

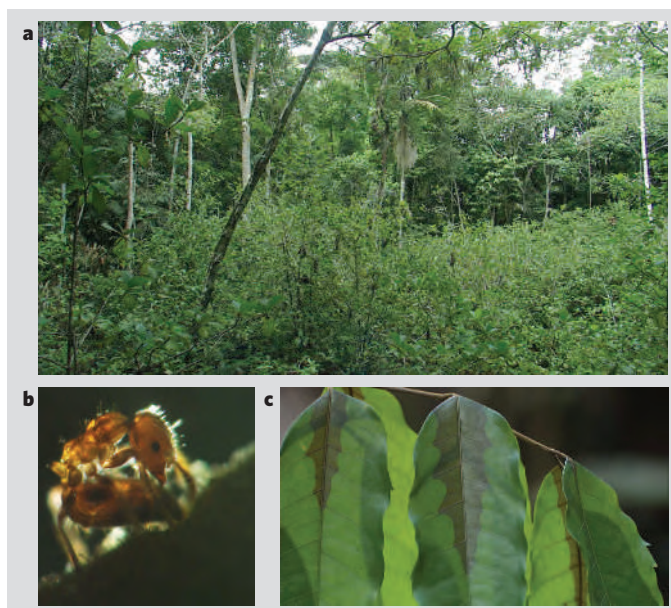
'Devil's gardens' are large stands of trees in the Amazonian rainforest that consist almost entirely of a single species, *Duroia hirsuta*^{1–5}, and, according to local legend, are cultivated by an evil forest spirit. Here we show that the ant *Myrmelachista schumanni*, which nests in *D. hirsuta* stems, creates devil's gardens by poisoning all plants except its host plants with formic acid. By killing these other plants, *M. schumanni* provides its colonies with abundant nest sites — a long-lasting benefit as colonies can live for 800 years.

M. schumanni lives in the hollow, swollen stems (domatia) of *D. hirsuta*, the tree species that dominates devil's gardens (Fig. 1a). Previous studies of the mutualism between *D. hirsuta* and *M. schumanni* indicated that devil's gardens result from allelopathy, which is the local inhibition of plant growth by another plant, by *D. hirsuta*^{2–5}. However, studies of a different ant–plant mutualism — between an unidentified species of *Myrmelachista* and the ant-plants *Tococa guianensis* and *Clidemia heterophylla* — indicated that *Myrmelachista* may create stands comprising only its host plants by using herbicide^{6,7}.

We did an ant-exclusion experiment to determine whether the selective killing of plants inside devil's gardens is due to the activity of *M. schumanni* workers or to allelopathy by *D. hirsuta*. We planted saplings of a common Amazonian tree, the cedar *Cedrela odorata*, inside and outside devil's gardens, and either excluded or did not exclude ants from the saplings (for methods, see supplementary information).

We found that the *M. schumanni* workers

Figure 1 | The ant *M. schumanni* creates devil's gardens by killing all plants other than its host tree, *D. hirsuta*. **a**, A devil's garden, or mono-specific stand of *D. hirsuta*, in the foreground contrasts with the species-rich rainforest in the background. **b**, A worker *M. schumanni* ant attacking a plant: the ant bites a small hole in the leaf tissue, inserts the tip of its abdomen into the hole and releases formic acid. **c**, Leaves develop necrosis along primary veins within hours of the attack.



M. DOHRN © BBC

promptly attacked the saplings in devil's gardens from which ants had not been excluded, injecting a poison into their leaves (Fig. 1b), which developed necrosis within 24 hours (Fig. 1c). Most of the leaflets on these saplings were lost within five days, and the proportion lost was significantly higher than on saplings from which ants were excluded (Fig. 2). We also found that ant-free *C. odorata* inside devil's gardens fared as well as *C. odorata* planted outside devil's gardens. These results show that devil's gardens are produced by *M. schumanni* workers, rather than by *D. hirsuta* allelopathy.

In a second experiment, we investigated

whether *M. schumanni* attacks only plants that are not its host plants and whether the ant uses domatia to recognize its host. We planted *C. odorata* saplings with and without artificial domatia and *D. hirsuta* saplings with and without domatia in devil's gardens. After 24 h, there was significant leaf necrosis on all *C. odorata* plants (mean area on plants with artificial domatia: 39.7 cm²; s.e., 26.4–55.6 cm²; on plants without artificial domatia: 14.2 cm²; s.e., 9.2–20.3 cm²), whereas there was no leaf necrosis at all on any *D. hirsuta* plants, irrespective of the presence of domatia (analysis of variance, $F_{3,20} = 57.03$, $P < 0.0001$). We conclude that *M. schumanni* attacks only non-host plants such as *C. odorata* and that it does not rely on the presence of domatia to discriminate between its hosts and other plant species.

Chemical analysis revealed that the poison glands of *M. schumanni* contain formic acid (0.43 ± 0.12 μl per worker); no other compounds were detected. Treatment of leaves with formic acid induced leaf necrosis on all the plants we tested. (For details, see supplementary information.) Many formicine ants produce formic acid: to our knowledge, this is the first record of an ant using formic acid as a herbicide — although it is known to have bactericidal and fungicidal properties⁸.

Devil's gardens covered 4.5% of our study plot and grew by $0.7 \pm 0.3\%$ per year. Using this growth rate, we estimate that the largest

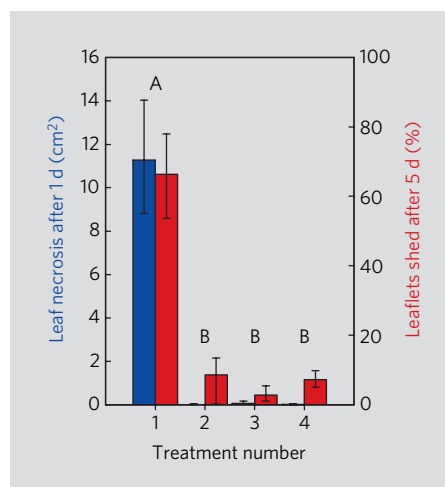


Figure 2 | *M. schumanni* ants, and not allelopathy, create devil's gardens. Saplings of the non-host plant *C. odorata* were subjected to different treatments: 1, planted inside a devil's garden, ants not excluded; 2, planted inside a devil's garden, ants excluded; 3, planted outside devil's gardens, ants not excluded; and 4, planted outside devil's gardens, ants excluded. Only saplings exposed to ants inside devil's gardens developed significant necrosis within one day (average $\pm$ s.e.; blue bars) and shed a significant percentage of their leaflets within five days (average $\pm$ s.e.; red bars). Multivariate analysis of variance results: Pillai trace, 0.88, $F_{6,72} = 9.41$, $P < 0.0001$. Analysis of variance (ANOVA) results (necrosis): $F_{3,36} = 52.78$, $P < 0.0001$. ANOVA results (leaflets shed): $F_{3,36} = 17.19$, $P < 0.0001$. Bars marked A are significantly different ($P < 0.001$) by Tukey post-hoc tests from bars marked B.

devil's garden in our plot, with 351 plants, is 807 years old (95% confidence interval, 446–4,234 years old). A devil's garden is tended by a single *M. schumanni* colony (our unpublished results) comprising as many as 3 million workers and 15,000 queens; the presence of many queens undoubtedly contributes to colony longevity.

The cultivation of devil's gardens by *M. schumanni* is an example of niche construction⁹. By killing plants of other species, the ant promotes the growth and establishment of *D. hirsuta*, thereby gaining more nest sites. A devil's garden begins when an *M. schumanni* queen colonizes a single *D. hirsuta* tree: over time, *D. hirsuta* saplings become estab-

lished within the vegetation-free area created by the ants, and the ant colony expands to occupy them. The devilry of *M. schumanni* today provides homes for ants in the future.

Megan E. Frederickson*, Michael J. Greene†, Deborah M. Gordon*

*Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA
e-mail: meganf@stanford.edu

†Department of Biology, University of Colorado at Denver and Health Sciences Center, Denver, Colorado 80217-3364, USA

1. Davidson, D. W. & McKey, D. *Trends Ecol. Evol.* **8**, 326–332 (1993).
2. Campbell, D. G., Richardson, P. M. & Rosas Jr, A. *Biochem. Syst. Ecol.* **17**, 403–407 (1989).

3. Page, J. E., Madriñán, S. & Towers, G. H. N. *Experientia* **50**, 840–842 (1994).
4. Aquino, R., De Tommasi, N., Tapia, M., Lauro, M. R. & Rastrelli, L. J. *Nat. Prod.* **62**, 560–562 (1999).
5. Pfannes, K. R. & Baier, A. C. *Rev. Biol. Trop.* **50**, 293–301 (2002).
6. Morawetz, W., Henzl, M. & Wallnöfer, B. *Biodivers. Conserv.* **1**, 19–33 (1992).
7. Renner, S. S. & Ricklefs, R. E. *Biotropica* **30**, 324–327 (1998).
8. Revis, H. C. & Waller, D. A. *Auk* **121**, 1262–1268 (2004).
9. Odling-Smee, F. J., Laland, K. N. & Feldman, M. W. *Niche Construction: The Neglected Process in Evolution* (Princeton Univ. Press, Princeton, 2003).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.
doi:10.1038/437495a

CLIMATE MODELLING

Northern Hemisphere circulation

Air pressure at sea level during winter has decreased over the Arctic and increased in the Northern Hemisphere subtropics in recent decades, a change that has been associated with 50% of the Eurasian winter warming observed over the past 30 years, with 60%

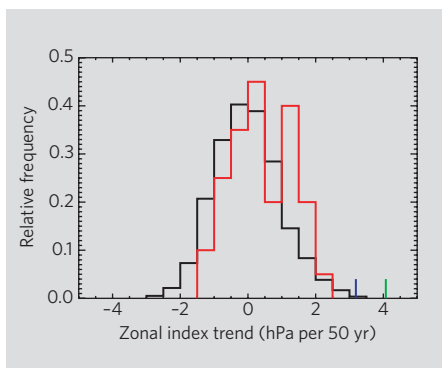


Figure 1 | Simulated and observed zonal index trends. The green line represents the 1955–2005 trend in the mean sea-level pressure in the Northern Hemisphere over December–February southwards of 45° N minus the mean sea-level pressure northwards of 45° N taken from the NCEP reanalysis⁴ and the blue line represents the same from the HadSLP2r data set (R. J. Allan and T. J. Ansell, manuscript submitted). The black line is a normalized histogram of 50-year zonal index trends in 3,903 overlapping segments of control simulations from nine coupled climate models obtained from the Intergovernmental Panel on Climate Change data archive of the Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory. The red line is a normalized histogram of 1955–2005 zonal index trends from 40 historical simulations of the nine models with anthropogenic and natural forcings. In cases where the historical simulations ended before February 2005, output from twenty-first century scenario integrations with only anthropogenic forcings was used for the missing years.

of the rainfall increase in Scotland and with 60% of the rainfall decrease in Spain¹. This trend is inconsistent with the simulated response to greenhouse-gas and sulphate-aerosol changes^{2,3}, but it has been proposed that other climate influences — such as ozone depletion — could account for the discrepancy³. Here I compare observed Northern Hemisphere sea-level pressure trends with those simulated in response to all the major human and natural climate influences in nine state-of-the-art coupled climate models over the past 50 years. I find that these models all underestimate the circulation trend. This inconsistency suggests that we cannot yet simulate changes in this important property of the climate system or accurately predict regional climate changes.

I derived a zonal index by subtracting the December–February mean sea-level pressure northwards of 45° N from the mean sea-level pressure between the Equator and 45° N using two data sets: the NCEP reanalysis⁴ and a newly completed data set of gridded observations known as HadSLP2r (R. J. Allan and T. J. Ansell, manuscript submitted). The trend in this zonal index over the period 1955–2005 is shown in Fig. 1 for both data sets. I compared the observed trend with output from nine coupled climate models from the Intergovernmental Panel on Climate Change data archive (UKMO-HadCM3, CCSM3, PCM, GFDL-CM2.0, GFDL-CM2.1, MIROC3.2(medres), MIROC3.2(hires), GISS-EH and GISS-ER).

To assess whether the observed trend could result from internal climate variability, I calculated equivalent zonal index trends in 3,903 overlapping 50-year segments of the models' control integrations (Fig. 1, black histogram). The NCEP zonal index trend exceeds all 50-year zonal index trends from the control simulations, and the HadSLP2r zonal index trend

exceeds all but one, indicating that the observed trend is inconsistent with simulated internal variability at the 5% significance level.

I assessed the possible role of external forcing by calculating zonal indices for the period 1955–2005 in historical simulations from the same climate models, which include greenhouse-gas, sulphate-aerosol, stratospheric-ozone, volcanic-aerosol and solar-irradiance changes. The resulting zonal index trends in each of the 40 ensemble members are shown by the red histogram in Fig. 1. Although the mean trend in these simulations is significantly positive (0.40 hPa per 50 yr), all the simulated trends are less than the observed trend, indicating that the simulated and observed trends are inconsistent at the 5% level, in contrast to previous findings based on the North Atlantic Oscillation index⁵.

Overall, I find that the observed Northern Hemisphere circulation trend is inconsistent both with simulated internal variability and with the simulated response to human and natural climate influences, although the mean simulated zonal index trend is consistent in sign with that observed. This is therefore an important aspect of large-scale climate change that these state-of-the-art climate models are unable to simulate; if we could understand and correct this bias, predictions of future regional climate change would be improved.

Nathan P. Gillett

Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK
e-mail: n.gillett@uea.ac.uk

1. Thompson, D. W. J., Wallace, J. M. & Hegerl, G. C. *J. Clim.* **13**, 1018–1036 (2000).
2. Gillett, N. P., Zwiers, F. W., Weaver, A. J. & Stott, P. A. *Nature* **422**, 292–294 (2003).
3. Osborn, T. J. *Clim. Dyn.* **22**, 605–623 (2004).
4. Kalnay, E. et al. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
5. Selten, F. M., Branstator, G. W., Dijkstra, H. A. & Kliphuis, M. *Geophys. Res. Lett.* **31**, L21205 (2004).

Competing financial interests: declared none.
doi:10.1038/437496a

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see Nature contents.

HUMAN COOPERATION

Second-order free-riding problem solved?

Arising from: K. Panchanathan & R. Boyd *Nature* 432, 499–502 (2004)

Panchanathan and Boyd¹ describe a model of indirect reciprocity in which mutual aid among cooperators can promote large-scale human cooperation without succumbing to a second-order free-riding problem² (whereby individuals receive but do not give aid). However, the model does not include second-order free riders as one of the possible behavioural types. Here I present a simplified version of their model to demonstrate how cooperation unravels if second-round defectors enter the population, and this shows that the free-riding problem remains unsolved.

Suppose a population shows two types of behaviour. In each period, 'cooperators' pay a cost C to contribute to a public good and receive a benefit B . 'Defectors' also receive benefit B but do not pay the cost. Assuming growth of each type in the population is proportional to mean pay-offs, how can we explain the emergence and persistence of cooperators?

Panchanathan and Boyd propose that a first-round public-goods game is followed by a second-round 'mutual-aid' game. In this game, each 'shunner' cooperates in the first round and then pays a cost c to generate a benefit b to one randomly chosen shunner, which yields a pay-off of $B - C + b - c$. Shunners can invade and dominate a population of defectors if $b - c > C$. In other words, shunners will prevail if they can create an in-group net surplus from mutual aid in the second round that exceeds the cost of cooperation in the first round. However, this is only possible if shunners can exclude second-round defectors from mutual aid. Although Panchanathan and Boyd include in their model individuals who receive a benefit without paying the cost in the first round (defectors), they do not consider such a type for the second round. Thus, rather than solving the second-order free-rider problem, their model merely assumes it away.

To see why, suppose that 'second-round defectors' enter the population. These individuals receive the same pay-off as shunners in the first round and are eligible to receive aid in the second round. However, they do not give aid to others. If we denote the proportion of shunners in the population by p , then the average pay-off for second-round defectors is $B - C + bp$ and the average pay-off to shunners changes to $B - C + bp - c$. If second-round defectors invade a population of shunners, then second-round cooperation collapses, pay-offs to second-round defectors fall, and eventually ordinary defectors can invade and dominate the population — leaving us back where we started.

Panchanathan and Boyd implicitly acknowledge this problem when they note that each individual must have perfect information about the cooperation and aid-giving histories of all other members of the population in order for mutual aid to be sustained. In other words, shunners must be able to recognize and exclude second-round defectors from receiving aid. They do incorporate an error term into their model, but they do not consider errors in which individuals mistakenly help a recipient of bad reputation during the mutual-aid game¹.

Suppose that e is the probability that shunners mistakenly aid a second-round defector or withhold aid from a shunner. Cooperation can only be maintained when the pay-off to shunners, $B - C + bp(1 - e) - c[p(1 - e) + e(1 - p)]$, is greater than the pay-off to second-round defectors, $B - C + ebp$, or

$$e < \frac{p(b - c)}{2p(b - c) + c}$$

This means a population of shunners ($p = 1$) is only evolutionarily stable if the error is sufficiently small

$$e < \frac{b - c}{2b - c}$$

In other words, cooperation will unravel if second-round defectors cannot be detected most of the time. In contrast, a population of second-round defectors ($p = 0$) is stable for any positive error rate and can resist invasion even when shunners are common. Thus, the emergence of shunners when they are rare cannot be explained by the authors' model¹.

Note that the simple model presented here raises a broader concern with all models of indirect reciprocity^{3–5} and related experimental results⁶. Previous work has already shown that indirect reciprocity is stable only when donors have very reliable information about the behavioural histories of all individuals in the population^{4,7}. But even this assumes that there is no evolutionary pressure on the reliability of information. If at least some acts of giving are not observable, then individuals may have an incentive to misrepresent their behavioural histories in order to secure the benefits of indirect reciprocity without paying the costs. Considering a human context in particular, what keeps these individuals from evolving deceptive behaviours that would reduce the reliability of information and allow them to benefit from aid without providing it?

James H. Fowler

Department of Political Science, University of California at Davis, One Shields Avenue, Davis, California 95616, USA
e-mail: jhfowler@ucdavis.edu

1. Panchanathan, K. & Boyd, R. *Nature* **432**, 499–502 (2004).
2. Fehr, E. *Nature* **432**, 449–450 (2004).
3. Nowak, M. A. & Sigmund, K. *Nature* **393**, 573–577 (1998).
4. Panchanathan, K. & Boyd, R. *J. Theor. Biol.* **224**, 115–126 (2003).
5. Alexander, R. D. *The Biology of Moral Systems* (de Gruyter, New York, 1987).
6. Milinski, M., Semmann, D. & Krambeck, H. J. *Nature* **415**, 424–426 (2002).
7. Nowak, M. A. & Sigmund, K. *J. Theor. Biol.* **194**, 561–574 (1998).

doi:10.1038/nature04201

HUMAN COOPERATION

Panchanathan & Boyd reply

Reply to: J. H. Fowler *Nature* 437, doi:10.1038/nature04201 (2005)

We have shown that, if a system of indirect reciprocity is stable, exclusion from that system could deter collective-action cheats¹. Unlike direct punishment^{2–5}, indirect punishers benefit by avoiding donation, obviating the second-order free-rider problem. Fowler claims⁶, however, that we assume away the second-order free-rider problem, and (by adding a new error term) argues that indirect-reciprocity defectors undermine cooperation.

We find three flaws in Fowler's argument. First, we do not assume away the second-order free-rider problem. In models with direct punishment^{2–5}, the public-goods phase represents

a first-order collective action, whereas the punishment phase represents a second-order collective action. Such models are vulnerable to the second-order free-rider problem because selection favours strategies that avoid costly punishment of public-goods cheats (that is, second-order free riders). In our model¹, the public-goods phase is followed by an indirect-reciprocity phase. Because public-goods cheats are punished through exclusion from indirect reciprocity, selection favours punishers (shunners) over non-punishers (cooperators). Whereas cooperators represent second-order free riders, indirect punishment

is not susceptible to the second-order free-rider problem.

Second, Fowler's second-round defectors represent analogues of defecting strategies considered in previous models of indirect reciprocity^{7,8}. If they can avoid detection, Fowler is correct: indirect reciprocity may not be stable^{7,8}. Our model¹ assumes indirect reciprocity is stable as a point of departure for analysing a different problem. To link Fowler's claim to previous results, we present a model combining a public-goods game and an iterated indirect reciprocity game^{7,8}, considering only shunners and second-round defectors. In Fowler's model, error about individuals' intentions and knowledge about others' past behaviour are conflated in a single term.

Disentangling these, we first consider implementation errors⁸, in which individuals defect when they meant to cooperate and cooperate when they meant to defect. We omitted unintentional cooperation error in our original model¹ as it has no qualitative effect on the results. We also consider ignorance about the past behaviour of partners.

Using this model, we find that shunners can resist invasion by rare second-round defectors if

$$\left(\frac{wq(1-e)}{1-wqe} \right) b > c \quad (1)$$

where q represents the probability of knowing the standing (either good or bad) of a ran-

domly selected partner, w represents the probability that the indirect reciprocity game persists for another interaction, b is the benefit of receiving mutual aid, c is the cost of providing mutual aid, and e is the probability that a shunner defects when intending to cooperate. (Note the opposite error term, intentional defection resulting in cooperation, is absent in the condition described in equation (1).) This means that individuals are ignorant about the behaviour history of their partner with probability $1 - q$ and the expected number of interactions in the indirect reciprocity stage is $1/(1 - w)$. The bracketed expression on the left of condition (1) represents the degree of assortment.

In order for cooperation to evolve in any model, altruists must be able to channel cooperation to other altruists and withhold it from defectors. In indirect reciprocity, assortment is generated through reputation^{7,8}. If reputations are sufficiently accurate, cooperation can be evolutionarily stable. For example, if we set $e = 0$, the shunner strategy is stable if $wqb > c$. In inclusive-fitness models⁹, assortment is generated through kin-biased interaction, leading to Hamilton's rule ($rb > c$, where r is the coefficient of relatedness). In reciprocal-altruism models¹⁰, the persistence of dyads generates assortment, leading to the expression $wb > c$.

Third, finding that shunners cannot invade a population of second-round defectors,

Fowler raises a broad concern with all models of indirect reciprocity. This is based on a misunderstanding: in systems of reciprocity, both direct¹⁰ and indirect^{7,8}, randomly paired altruists can only increase in frequency if their initial frequency exceeds a threshold. Below this threshold, altruists too often interact with defectors and cooperation collapses. Fowler's finding that shunners cannot invade second-round defectors is therefore unremarkable. If kin-biased interactions are assumed, previous models of reciprocity have shown that altruists can increase when rare^{1,8,10}.

Karthik Panchanathan, Rob Boyd

Center for Behavior, Evolution, and Culture and Department of Anthropology, University of California, Los Angeles, California 90095, USA
e-mail: buddha@ucla.edu

1. Panchanathan, K. & Boyd, R. *Nature* **432**, 499–502 (2004).
2. Boyd, R. & Richerson, P. *Ethol. Sociobiol.* **13**, 171–195 (1992).
3. Gintis, H. *J. Theor. Biol.* **206**, 169–179 (2000).
4. Henrich, J. & Boyd, R. *J. Theor. Biol.* **208**, 79–89 (2001).
5. Boyd, R., Gintis, H., Bowles, S. & Richerson, P. *Proc. Natl Acad. Sci. USA* **100**, 3531–3535 (2003).
6. Fowler, J. H. *Nature* doi:10.1038/nature04201 (2005).
7. Nowak, M. & Sigmund, K. *J. Theor. Biol.* **194**, 561–574 (1998).
8. Panchanathan, K. & Boyd, R. *J. Theor. Biol.* **224**, 115–126 (2003).
9. Hamilton, W. D. *J. Theor. Biol.* **7**, 1–52 (1964).
10. Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390–1396 (1981).

doi:10.1038/nature04202

REVIEWS

Pathophysiological consequences of VEGF-induced vascular permeability

Sara M. Weis¹ & David A. Cheresh¹

Although vascular endothelial growth factor (VEGF) induces angiogenesis, it also disrupts vascular barrier function in diseased tissues. Accordingly, VEGF expression in cancer and ischaemic disease has unexpected pathophysiological consequences. By uncoupling endothelial cell-cell junctions VEGF causes vascular permeability and oedema, resulting in extensive injury to ischaemic tissues after stroke or myocardial infarction. In cancer, VEGF-mediated disruption of the vascular barrier may potentiate tumour cell extravasation, leading to widespread metastatic disease. Therefore, by blocking the vascular permeability promoting effects of VEGF it may be feasible to reduce tissue injury after ischaemic disease and minimize the invasive properties of circulating tumour cells.

The VEGF family includes VEGF-A, placenta growth factor, VEGF-B, VEGF-C and VEGF-D. VEGF-A (hereafter referred to as VEGF) is a key regulator of normal and pathological blood vessel growth. VEGF was first described as a vascular permeability factor (VPF) by Dvorak and colleagues¹. In the ensuing 20 years, a large body of research has evolved characterizing the molecular pathways that mediate the distinct biological activities of VEGF once bound to its cell-surface receptors VEGFR1 (also known as Flt-1), VEGFR2 (also known as Flk-1/KDR) and neuropilin-1/2 (NP1/NP2)². Previous studies^{3,4} have shown that targeted deletion of VEGF in mice leads to abnormal vascular development and embryonic lethality, emphasizing the critical role for VEGF in blood vessel development. A recent review⁵ portrays the history of VEGF-A/VPF and the discoveries leading to the development and clinical approval of the anti-VEGF antibody bevacizumab to target tumour angiogenesis.

The VEGF family of growth factors is unique, as it comprises the only angiogenic factor that also potently induces vascular leak. Although VEGF-induced angiogenesis is often accompanied by a vascular permeability response, VEGF-induced vascular leak is not required for angiogenesis⁶. VEGF-mediated vascular permeability leads to accumulation of a fibrin barrier around tumours and wounds⁷, which may limit the malignant properties of these tissues. However, it has become clear that the vascular permeability-producing effects of VEGF may have profound negative consequences in ischaemic disease by augmenting the level of infarcted tissue^{8,9}. Aside from the pathological effects of VEGF-mediated vascular permeability, it is unclear how this process contributes to normal physiology after development. In addition, potential strategies to treat cancer and ischaemic disease may benefit from a better understanding of VEGF-induced vascular leak. Thus, it is important to understand the molecular mechanisms leading to VEGF-induced permeability in both physiological and pathological states. Although many studies have focused on the role that VEGF plays as an angiogenic factor, this review will primarily discuss the vascular permeability-promoting effects of VEGF and the effects of hyper-permeability on the pathophysiology of disease.

VEGF induces vascular leak

Defining vascular permeability. Vascular permeability can be simply

defined as the movement of fluids and molecules between the vascular and extravascular compartments. Interestingly, vascular permeability can be regulated by VEGF as well as a wide array of inflammatory mediators. However, experimental evidence suggests that the vascular leak induced by VEGF and inflammatory mediators occurs via distinct molecular processes¹⁰. In any event, the process of vascular leak is quite complex, and depends on a number of variables (see Fig. 1), including the physical properties of the fluid or molecule being transported (size, charge, configuration), gradients between compartments (pressure or concentration) and the mode of transport (through channels or vesicles within an individual cell or through interendothelial junctions between adjacent cells). Very little macromolecular transport may occur despite high permeability if pressure gradients limit extravasation¹¹. Molecules of various sizes may extravasate by distinct means, with different kinetics, and with varying degrees of clearance by lymphatics¹².

Mechanisms of vascular leak. Ultrastructural studies have offered insight into how endothelial cells respond to VEGF during the vascular permeability response. Endothelial cells exposed to VEGF allow passage of particles of different sizes by a variety of physical mechanisms (Fig. 1). Roberts and Palade¹³ discovered that VEGF can induce the formation of fenestrations, or small pores, in the thin endothelial cell cytoplasm, allowing leak of small solutes. VEGF also induces formation of caveolae¹⁴, small plasmalemmal invaginations, which allow vesicular transport of small proteins through the cytoplasm of a single endothelial cell. Feng *et al.*¹⁵ coined the term vesiculo-vacuolar organelles (VVOs) to describe the fusion of vesicles to form a channel through the endothelial cytoplasm. These mechanisms may provide a physical pathway for very small solutes to pass through a single endothelial cell, but the leak of larger proteins probably occurs at the intercellular junctions between adjacent endothelial cells. One study investigated the functional limits of transvascular transport, and defined pore cutoff sizes ranging from 200 nm to 1.2 μ m in vessels associated with subcutaneous tumours¹⁶. Furthermore, the size and number of these junctional gaps and pores are affected by VEGF in a dose-dependent manner that varies between vascular beds, suggesting different dose thresholds for VEGF in different tissues or tumour types¹⁷. These studies suggest that precise levels of exogenous VEGF could be used to manipulate the vascular barrier and enhance delivery of anti-tumour agents.

¹Department of Pathology and Moores UCSD Cancer Center, University of California, San Diego, La Jolla, California 92093-0803, USA.

When endothelial cell–cell junctional integrity is disrupted *in vivo*, the consequence is leak of serum/plasma proteins and circulating cells from the blood vessels (Fig. 2). Oedema is the accumulation of fluids in the extracellular space, and is associated with deleterious swelling and increased interstitial pressure. Sites of vascular leak where underlying basement membrane is exposed can attract platelets⁹ or even tumour cells¹⁸ to a potential site for adhesion or transendothelial migration. Platelets attracted to these sites of disruption bind to von Willebrand factor, fibronectin or underlying collagen, leading to their activation. These activated platelets often deposit high levels of VEGF¹⁹, thus creating more leak and attracting more platelets. The aggregating platelets attracted to the initial exposed gap between endothelial cells may quickly occlude a small vessel and cause localized ischaemia, which may further induce VEGF expression and endothelial gaps.

Experimental evaluation of permeability. A number of methods have been used to evaluate the process of vascular permeability in response to a variety of permeability-inducing agents (Table 1). Each model has advantages and limitations, and ultimately yields an ‘indicator of permeability’ dependent upon the physical configuration of the experiment. The fundamental differences between the various *in vitro* assays are evident in their ability to detect the earliest significant response to a particular permeability factor. For example, the vascular leak response to thrombin ranges from 5 min to 2 h (refs 20, 21), whereas VEGF is from 2 to 6 h (refs 22, 23). These *in vitro* models are limited by the fact that permeability is typically evaluated on a single cell type, in a non-physiological two-dimensional conformation, without vascular flow or pressure, and in the absence of circulating cells. In addition, endothelial cells isolated from different vascular beds exhibit varying levels of responsiveness to VEGF²⁴. Nevertheless, these models provide an opportunity to study time- and dose-dependent permeability responses in cells subjected to pharmacological inhibitors, gene silencing, or expressing genetic mutations.

The most commonly used *in vivo* assay to evaluate vascular

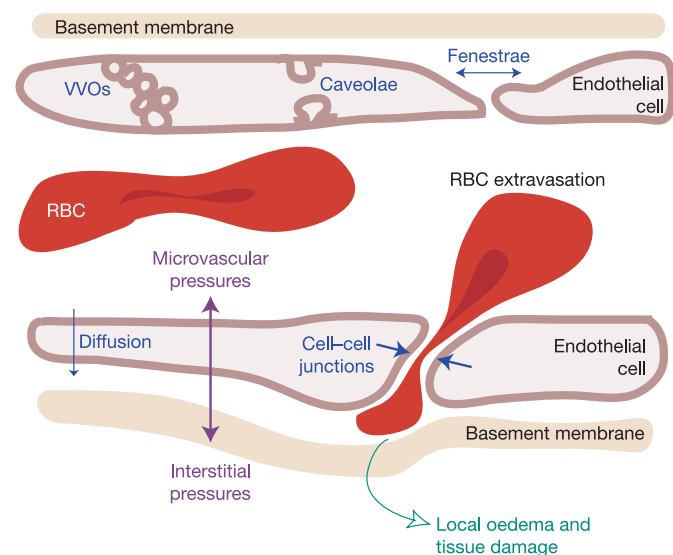


Figure 1 | Structural determinants of VEGF-induced leak. VEGF can induce leak of various-sized solutes via several routes through the endothelium, including by diffusion, interendothelial junctions, fenestrae, caveolae or VVOs. Hyperpermeability ultimately depends upon relative pressure gradients between the vascular and extravascular compartments, and thus a hyperpermeable vessel may not be ‘leaky’ if interstitial pressure is high relative to microvascular pressure. Normalization of hyperpermeable tumour-associated vessels (with bevacizumab, for example) is thought to improve drug delivery by facilitating a more homogeneous distribution of chemotherapy agents.

permeability *in vivo* is the Miles assay²⁵ (Table 1), which is often used to quantify vascular leak in the skin or the ear, and more recently in the lung²⁶. The Miles assay may be inappropriate to measure permeability changes in response to VEGF, because the measured solute flux is influenced by diffusive permeability, surface area, hydraulic conductivity, capillary pressure and interstitial pressure²⁷. Unfortunately, VEGF is likely to increase solute flux by changing most of these parameters, possibly independently of its direct effect on permeability.

Microscopy techniques have commonly been used to localize extravasation of macromolecular tracers of different sizes, which are labelled with heavy metals, radioisotopes, fluorescence or horseradish peroxidase. These techniques use intravital microscopy to monitor leak from single, isolated vessels or intact vascular beds from exteriorized thin tissues such as mesentery, or via window preparations such as the dorsal skin fold model. These methods have been used to illustrate vascular leak and oedema in the brain¹⁰, inflammation²⁸ and tumour-associated blood vessels²⁹. In particular, transmission electron microscopy has been used to evaluate endothelial barrier properties and extravasation of ferritin in blood vessels after VEGF administration¹⁵. Similarly, we have observed extravasation of red blood cells from coronary capillaries after myocardial infarction or systemic VEGF administration⁹ (Fig. 2).

VEGF receptor-mediated signalling pathways

Contribution of VEGF receptors. VEGF-mediated intracellular signals mediate the distinct activities of VEGF, including cell survival, proliferation, migration, vascular permeability and angiogenesis.

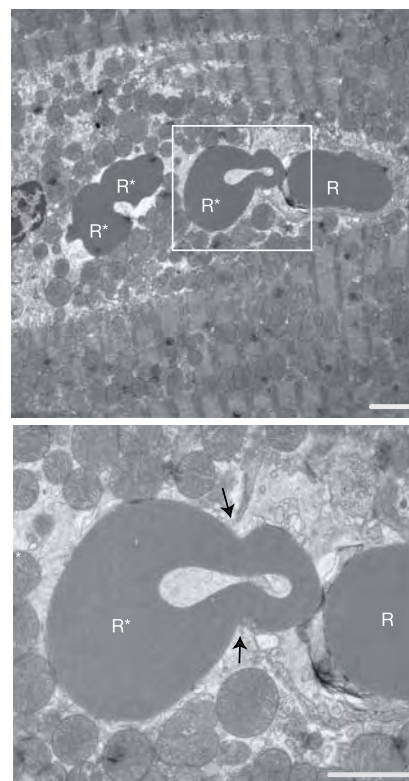


Figure 2 | VEGF induces vascular leak. Myocardial infarction or intravenous injection of VEGF induces vascular leak, observed here by transmission electron microscopy of a red blood cell in the process of extravasation from a capillary in a mouse heart through a gap between adjacent endothelial cells (arrows). Adjacent to this site of leak are more extravasated red blood cells and evidence of interstitial oedema and tissue injury. The bottom panel is a higher-magnification image of the boxed area in the top panel. R, red blood cell; R*, red blood cell outside the blood vessel. Scale bar, 2 μ m. Image adapted from ref. 9.

Table 1 | Methods to evaluate vascular permeability *in vitro* and *in vivo*

Model type	Description	Advantages	Limitations
<i>In vitro</i> models			
Boyden chamber/Transwell assay	Confluent endothelial cell monolayer seeded onto porous filter in upper chamber. Permeability measured as accumulation of tracer in lower chamber.	Multi-well format. Can visualize cells using microscopy and immunohistochemistry.	Relies on absolute confluence of monolayer, limited sampling frequency.
Transendothelial resistance or impedance	Permeability measured as changes in resistance or impedance across confluent cell monolayer.	Constant data acquisition. Can measure rapid and reversible responses.	Measurement may represent cell-cell and cell-matrix interactions.
Microcarrier beads	Cells grown to confluency on microcarrier beads. Permeability measured as ability of tracers to pass through cell monolayer into bead matrix or through cell column.	Can measure very rapid changes.	May have heterogeneous confluency between microcarrier beads.
<i>In vivo</i> models			
Miles assay	Measures permeability as extravasation of tracer from blood vessels in the skin induced by local injection of a permeability-inducing agent.	Can test several doses or agents in adjacent locations in same animal. Quantification of labelled tracers by extraction.	Measured solute flux may be influenced by haemodynamic parameters. Possible variation due to vessel morphology in skin.
Leak from single vessel	Microelectrodes are used to measure concentration gradients, or microscopy used to visualize movement of injected tracer.	Can use light, fluorescence, or electron microscopy to evaluate structural basis for leak.	Requires tissue manipulation. Individual vessels may not be representative of vascular beds.
Intravital microscopy of vascular beds	Extravasation of tracer measured or visualized from exteriorized thin tissues or window preparations.	Can identify sites of leak and quantify extent of extravasated tracer.	Requires some tissue manipulation. Difficult to use for long-term studies.

Although VEGF receptors were originally thought to be expressed solely on endothelial cells, the tissue distribution of VEGF receptors now includes vascular smooth muscle cells³⁰, osteoblasts³¹, cardiac myocytes³², myofibroblasts³³, neurons³⁴ and various tumour cells³⁵. In endothelial cells, VEGFR2 localization may determine specific signalling events after VEGF stimulation. For example, VEGFR2 associated with cadherins at cell–cell junctions⁹ may facilitate the vascular permeability response, whereas VEGFR2 associated with α v integrins³⁶ (α v β 3 and α v β 5) at the cell–matrix interface may influence permeability as well as angiogenesis (Fig. 3). Downstream signalling by VEGF is quite diverse, and includes transduction molecules such as phosphatidylinositol-3-OH kinase (PI(3)K)/Akt, Ras/Raf/MEK/Erk, Src and phospholipase C γ (PLC- γ)/endothelial nitric oxide synthase (eNOS). A wide array of factors has been implicated in the VEGF-induced permeability response, some of which are discussed below and illustrated in Fig. 3.

Cooperation with α v integrins. Growth factor receptor signalling is known to coordinate with that of integrins. There is a growing body of evidence that VEGF receptors and integrins are physically and functionally linked and may influence communication between cells and their underlying extracellular matrix. VEGF signalling *in vivo* seems to depend on integrin α v β 5, because anti- α v β 5, and to a lesser extent anti- α v β 3, blocks the angiogenic response induced by VEGF³⁷. In contrast, anti- α v β 3 but not anti- α v β 5 blocks basic fibroblast growth factor (bFGF)-mediated angiogenesis³⁷. However, distinct downstream effects of VEGF may be differentially regulated by these integrins, as β 5-null mice show little or no vascular leak in response to VEGF³⁸, whereas β 3-null mice show an increased vascular leak response³⁹. Together, these findings emphasize that VEGF signalling depends on cooperation between growth factor receptors and particular integrins, and highlights the specificity of these relationships in terms of downstream activities. There is biochemical evidence that the extracellular domain of α v β 3 integrin associates with VEGFR2 in cultured cells and influences cell migration and proliferation⁴⁰. It is possible that the relationship between β 3 and VEGFR2 provides negative regulation of VEGF signalling. In fact, mice lacking β 3 integrin exhibit increased VEGFR2 expression and are more responsive to VEGF in terms of angiogenesis⁴¹ and vascular leak³⁹, which may explain the increased tumour growth and metastasis observed in these mice³⁶.

VEGF modulates cell–cell junctions. VEGF receptors may also have a critical role in cell–cell adhesion and communication through their direct interaction with cell–cell adhesion molecules (Fig. 3). In endothelial cells, VEGFR2 associates with vascular endothelial (VE)-cadherin and provides regulation of cell–cell junctional integrity⁹, and represents a mechanism for localized disruption by VEGF. In normal mouse cardiac blood vessels we have identified a pre-formed complex between VEGFR2 and VE-cadherin that rapidly and transiently dissociates in response to VEGF stimulation⁹. After ischaemic injury of the heart or brain, gaps can be seen between the endothelial cells lining small blood vessels in these tissues⁹ (Fig. 2).

The integrity of endothelial cell–cell junctions and vascular barrier function is regulated by a series of adhesion molecules that make up tight, gap and adherens junctions⁴² (Fig. 3). These gaps must be large enough for the extravasation of red blood cells, and probably involve disruption of some or all of these junctional proteins. VEGF blocks gap junctional communication between adjacent endothelial cells by altering connexin-43 phosphorylation via VEGFR2 and Src kinase activation⁴³. VEGF also disrupts tight junctions by altering phosphorylation of zonula occludens-1 (ZO-1) and occludin⁴⁴ through a Src-dependent pathway⁴⁵. Several adherens junction proteins (VE-cadherin, β -catenin, γ -catenin/plakoglobin and p120-catenin) become tyrosine phosphorylated downstream of VEGFR2 after VEGF stimulation²², leading to the loosening of cell–cell contacts between endothelial cells *in vitro*. *In vivo*, disruption of VE-cadherin with a function-blocking antibody leads to vascular leak and haemorrhage⁴⁶. We have recently demonstrated that Src or Yes kinase activity is required for VEGF-induced phosphorylation of these adherens junction proteins, which leads to vascular leak *in vivo*, as mice deficient in either of these kinases show a complete loss of VEGF-mediated vascular permeability⁹. Previous *in vitro* studies have shown that pharmacological Src inhibition can block permeability and VE-cadherin tyrosine phosphorylation induced by hydrogen peroxide⁴⁷ or TNF- α ⁴⁸. Recently, a role for Src-dependent VE-cadherin tyrosine phosphorylation downstream of VEGF in blood vessels and cultured endothelial cells has been confirmed⁴⁹.

This requirement for Src activity may be a phenomenon shared by other growth factor–cadherin pairs, because N-cadherin–bFGF-R⁵⁰ and E-cadherin–epidermal growth factor receptor (EGFR)⁵¹ represent similar signalling complexes that transduce signals upon

stimulation by their cognate growth factors. A similar signalling pathway may regulate the epithelial–mesenchymal transition in tumour cells, in which EGF-R-mediated Src kinase activity initiates the dissociation of E-cadherin-mediated cell–cell adhesion⁵². This signalling mechanism might then allow migration of individual tumour cells from a primary tumour during local invasion or distant metastasis.

Alteration of cadherin-mediated cell–cell adhesion is often correlated with changes in tyrosine phosphorylation of both cadherins and catenins, which alters barrier function by disrupting the bridge of proteins linking adherens junctions to the actin cytoskeleton. Phosphorylation of a single tyrosine residue in β -catenin (Y654)⁵³ or p120 (Y217)⁵⁴ regulates their interaction with E-cadherin. Classical cadherins, including VE-cadherin, have several tyrosine phosphorylation sites within their cytoplasmic domain. In the N-cadherin cytoplasmic tail, mutation of Y833 decreases phosphorylation of N-cadherin by Src⁵⁵. The presence of protein tyrosine phosphatases (PTPs) at the adherens junction may also regulate phosphorylation status of cadherins and catenins (Fig. 3). Indeed, PTP1B⁵⁶, PTP μ ⁵⁷, VE-PTP⁵⁸, DEP-1 (ref. 59) and SHP-2 (ref. 60) have been shown to influence cadherin–catenin interaction and intercellular adhesion. In fact, the latter two PTPs participate in signalling complexes with VEGFR2 (refs 59, 61).

A variety of other signalling pathways initiated by VEGF and VEGFR2 may contribute to vascular leak (Fig. 3). For example, inhibitors of eNOS, PLC- γ , or protein kinase C can each block VEGF-induced permeability⁶². In addition, a sustained VEGF-induced permeability response is mediated by activation of the urokinase plasminogen activator and receptor system (uPA/uPAR)⁶³. In the absence of VEGF ligand binding, VEGFR2 activation by shear stress is sufficient to induce a permeability response that requires integrin signalling and results in rearrangement of adherens

junction proteins^{64,65}. VEGF activates p21-activated kinase (PAK), which locates to cell–cell junctions and is required for myosin light chain phosphorylation and cell contractility⁶⁶. Even focal adhesion kinase (FAK), which has a critical role in integrin-mediated signalling, is required for VEGF-induced hyperpermeability⁶⁷. Although signal transduction from kinases, phosphatases, cell–cell or cell–matrix adhesion molecules, and physical forces are each required to induce vascular leak from VEGF stimulation, it remains poorly understood how these apparently disparate signalling pathways converge to create hyperpermeability in response to VEGF.

Effects of hyperpermeability in pathological states

VEGF expression during ischaemic disease. VEGF is induced under a variety of circumstances in tissues. Hypoxic conditions result in the stimulation of hypoxia-inducible factors (HIF), which translocate to the nucleus to bind hypoxia response elements (HRE) in target genes. Hypoxic stimuli during development appear to influence VEGF expression and thereby regulate vasculogenesis or angiogenesis. In embryonic tissues, hypoxia induces VEGF expression and endothelial cell proliferation and migration, whereas hyperoxia produces an opposite response⁶⁸. During ischaemic injury in mature tissues, the induction of VEGF may occur through similar HIF-dependent pathways, and is associated with positive remodelling and re-vascularization of damaged tissues. In the brain after stroke, VEGF expression appears within 6 h in response to hypoxia along the ischaemic border zone⁶⁹. Although it seems to be constructive in the long term—because the expression of VEGF, VEGFR2 and HIF-1 in this hypoxic region correlates with the strong increase in the number of newly formed vessels⁶⁹—this response may contribute to deleterious oedema formation during the acute stage⁸.

The response to ischaemia is similar in the heart after myocardial infarction, as observed in the brain after stroke, wherein VEGF and

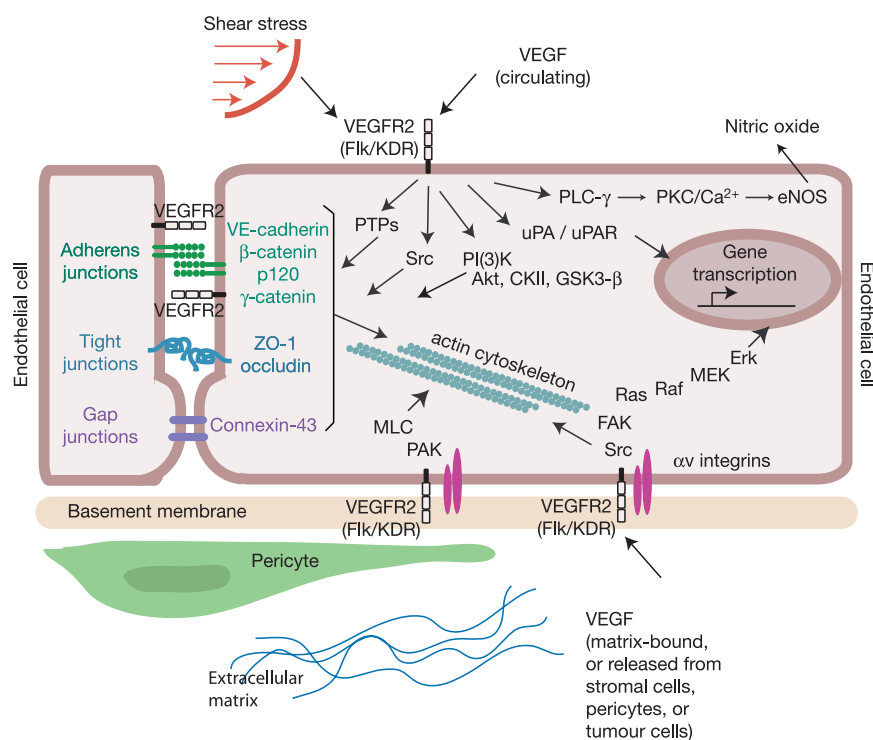


Figure 3 | Signalling pathways implicated in VEGF-induced vascular permeability. Plasma or serum VEGF released from circulating cells can transmit signals from the luminal surface of the vascular endothelium. Alternatively, VEGF bound to the extracellular matrix or released from pericytes, stromal cells, or tumour cells from the abluminal surface can initiate signalling pathways, leading to vascular leak. VEGF-induced

disruption of cell–cell adherens, tight and gap junctions often involves activation of Src family kinases and/or PTPs. VEGFR2 localization to these cell–cell junctions may facilitate a specific, localized response. VEGFR2 also interacts with α v integrins, which transmit cell–matrix signals and link focal adhesions to the actin cytoskeleton.

VEGF receptor levels increase in the infarct border zone after the acute ischaemic event⁷⁰ (Fig. 4). VEGF-mediated vascular permeability contributes to damaging oedema formation within several hours⁹, but also promotes re-vascularization and long-term cell survival. Ischaemic preconditioning preceding acute myocardial infarction is also associated with increased VEGF expression, and provides cardioprotection attributable to increased blood vessel density⁷¹ and reduced endothelial cell apoptosis⁷². VEGF gene delivery has been investigated as a therapeutic option to improve outcome after myocardial infarction. Intramyocardial injection of naked DNA encoding HIF-1 α and VEGF₁₆₅ enhances angiogenesis and reduces infarct size in pre-clinical acute myocardial infarction models⁷³. Endothelial progenitor cell mobilization or transplantation into ischaemic myocardium leads to increased VEGF expression and shows promise for enhanced cytoprotection and angiogenesis⁷⁴. In addition, there is a synergistic effect of bone marrow mobilization and VEGFR2 gene therapy in myocardial ischaemia⁷⁵. However, VEGF expression must be controlled as it could lead to haemangioma-like intravascular tumours⁷⁶ after myocardial infarction.

Within hours after an acute myocardial infarction, VEGF induces gaps between adjacent endothelial cells lining the microvessels in the heart, allowing leak of serum, plasma proteins and red blood cells into the myocardium (Fig. 2), which exacerbates tissue injury and increases infarct size⁹. In fact, this leak response produces a localized tissue injury as seen in Fig. 2. Blocking VEGF activity within the first 6 h after ischaemic stroke reduces oedema and improves outcome⁸.

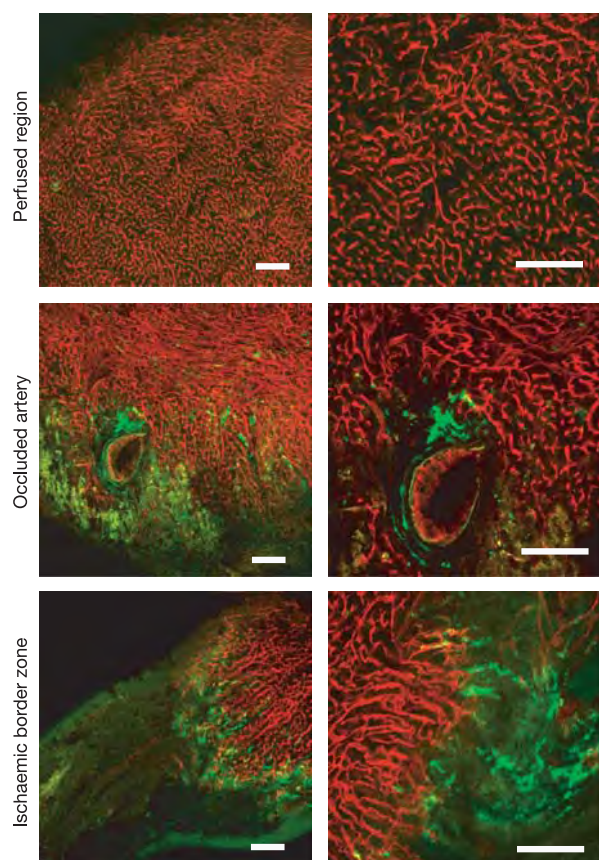


Figure 4 | VEGF expression after myocardial infarction. VEGF expression is induced by hypoxic stimuli after myocardial infarction, primarily along the border between perfused and non-perfused tissue. VEGF expression by multiple cell types along this ischaemic border zone is clearly visible 48 h after permanent occlusion of the left anterior descending coronary artery in mice expressing green fluorescent protein (green) under the VEGF promoter. Intravenous perfusion with rhodamine-tagged lectin labels endothelial cells in perfused vessels (red). Scale bars, 150 μ m.

VEGF-induced vascular permeability can be suppressed with blockade of c-Src^{9,71}, which we have shown to be necessary for VEGF signalling leading to vascular leak. Accordingly, genetic or pharmacological Src blockade after either stroke or myocardial infarction prevents oedema, reduces infarct size and preserves ventricular function without affecting VEGF expression^{9,10}. On the basis of this concept, a new inhibitor of VEGF-induced vascular leak is in phase II/II clinical trials for patients with acute myocardial infarction (J. Doukas, personal communication). This therapeutic strategy is the first to test whether blocking the early permeability-inducing effects of VEGF can preserve ischaemic tissue, ultimately leading to improved disease outcome and survival.

VEGF expression in cancer. VEGF was originally discovered as a VPF, a protein secreted by tumour cells that potently stimulates ascites formation and vascular leak¹. The phenomenon of vascular leak proximal to a tumour may promote extravasation of fibrin and attract the cellular machinery and materials required to construct new blood vessels to feed the primary tumour. This vascular leak probably contributes to the elevated interstitial pressure common in solid tumours, which, along with the abnormal properties of tumour-associated vessels, can compromise drug delivery to the tumour⁷⁷. In tumour-bearing animals⁷⁷ as well as rectal carcinoma patients⁷⁸, blockade of VEGFR2 induces tumour-associated vessel 'normalization', which improves outcome by decreasing the interstitial pressure and improving the function of the vascular network for delivery of cytotoxic therapies.

Elevated levels of circulating VEGF provide a predictive measure of the progression of cancer and metastasis in certain cancers⁷⁹, although inconsistencies exist with the collection and interpretation of circulating VEGF levels⁸⁰. Dysregulated pathological angiogenesis and leaky blood vessels are common near tumours, and anti-VEGF treatment reduces tumour vascular permeability, vessel diameter, and can lead to tumour regression⁸¹. Recent clinical data suggest that although bevacizumab (Avastin) therapy alone has only modest effects on human tumours, it can significantly increase survival of patients with metastatic colorectal cancer when combined with chemotherapy⁸². Emerging pre-clinical^{77,83} and clinical⁷⁸ data indicate that the benefit of combination therapy may be due to normalization of tumour vessels by anti-VEGF antibodies. This normalization effect includes a decrease in vascular permeability and concomitant decrease in interstitial pressure, resulting in increased drug delivery and oxygenation. In fact, VEGF blockade can also normalize pericyte and basement membrane coverage of tumour-associated blood vessels⁸³. These apparent counter-intuitive findings and their cellular and molecular underpinnings are reviewed in a recent paper⁸⁴.

Tumour suppressor genes represent a natural defence against invasive tumour cells, which could be due, in part, to their regulation of VEGF production. For example, the tumour suppressor p53 downregulates VEGF, and its presence is sufficient to prevent v-Src-induced increases in VEGF gene transcription⁸⁵. Loss or mutation of the tumour suppressor gene von Hippel-Landau (VHL) is associated with increased VEGF and VEGF receptor expression and activity in some cancers⁸⁶. Thus, strategies to target oncogenes and/or protect tumour suppressor genes may provide benefit by limiting tumour-induced VEGF production. This would provide protection both by hindering tumour cell invasive properties as well as reducing the host angiogenic and permeability responses to the tumour cells. In fact, herceptin, which targets EGFR (also known as Her2 or neu), has been shown to normalize and lower vascular permeability of tumour-associated vasculature⁸⁷.

We have recently shown that VEGF-expressing tumour cells in the circulation appear to extend protrusions towards endothelial cell junctions and disrupt cell-cell contact to enable their extravasation during the metastatic cascade¹⁸. Importantly, mice deficient in Src or Yes show no vascular leak response to VEGF and consequently are resistant to tumour cell extravasation and metastasis¹⁸. Although

vascular leak may augment the angiogenic response, VEGF-induced vascular permeability is not required for VEGF-induced angiogenesis, because Src-deficient mice show a normal angiogenic response to VEGF⁶.

Consequences of vascular leak during disease. Breakdown of the vascular barrier probably has a critical role in cancer, as VEGF-induced platelet aggregates could certainly impede blood flow in tumour-associated blood vessels. Alternatively, extensive high vascular pressure could couple vascular and interstitial compartments near a tumour, causing flow stasis without complete vessel occlusion. Previous studies have examined the contribution of vascular leak, vessel compliance and interstitial fluid pressure on pressure–flow relationships in tumour vasculature⁸⁸. These findings suggest that the microcirculation surrounding a tumour can be markedly affected by vascular compliance and leakiness, and that as these factors increase, blood is diverted away from the centre of a tumour towards its periphery. Thus, drug delivery to a tumour may be hindered by vascular leak, and therapeutic agents would probably accumulate in the periphery of a tumour without penetrating its core.

The ability to modulate VEGF expression in a controlled micro-environment may provide a method to manipulate the vascular barrier. In some situations, increasing vascular permeability during drug delivery may allow for location-specific targeting of higher tolerated doses of compounds, which could be delivered via dermal patches or nasal spray. As discussed above, reducing tumour vessel leakiness might enhance tissue oxygenation and therapeutic drug delivery. In fact, the ability of VEGFR2 blockade to normalize tumour vasculature and improve drug delivery seems to involve decreased vascular leakiness along with increased pericyte coverage of tumour-associated blood vessels and a thickening of the vascular basement membrane⁸³. In this case, VEGFR2 blockade may indirectly affect pericyte and extracellular matrix compartments via angiopoietin-1 and MMP activation⁸³. It should be emphasized that the host cells are critical in tumour-associated angiogenesis, as VEGF may be secreted from tumour and host cells to affect vascular hyperpermeability in the local environment⁸⁹.

In addition to its beneficial influence on ischaemic tissues, blocking VEGF-induced vascular leak may be useful in the treatment of spinal cord injury⁹⁰, pulmonary oedema due to acute lung injury or mechanical ventilation⁹¹, cerebral oedema associated with acute liver failure⁹², and retinal oedema associated with diabetes⁹³. Oedema due to VEGF seems to be biochemically distinct from that induced by inflammation. This is based on the finding that Src-deficient mice show no VEGF-mediated oedema, yet do show a normal oedema response to inflammatory mediators⁶. However, disruption of cell monolayer integrity or the vascular barrier *in vivo* can be attributed to a wide range of agents other than VEGF, including thrombin and histamine. It will be important to understand how the various permeability-inducing agents differ in their action, and what makes VEGF-induced permeability unique at the biochemical and molecular levels. In some circumstances, it may be preferable to block permeability induced by both VEGF and inflammatory pathways. To this end, recent studies suggest that blocking PI(3)K activity may interfere with both pathways of vascular leak and thereby provide significant benefit in models of myocardial infarction.

Unexpected roles for VEGF. VEGF expression is linked to a growing collection of maladies and remedies. For example, nicotine is known to increase the risk of vascular disease, cancer and metastasis, and has recently been shown to cause a significant increase in endothelial cell VEGF expression⁹⁴. The carcinogenic effects of second-hand smoke are correlated with increased VEGF expression and plasma VEGF levels⁹⁵. Some alternative remedies may provide their reputed benefit by inadvertently influencing VEGF signalling pathways. For example, a compound derived from honeybee propolis reduces tumour cell expression of VEGF, thereby limiting tumour angiogenesis, invasion and metastasis⁹⁶. Green tea catechins suppress VE-cadherin tyrosine phosphorylation and inhibit Akt activation during VEGF-induced

angiogenesis⁹⁷. Resveratrol, a phytoestrogen found in medicinal plants thought to provide some of the cardioprotective effects of red wine ('the French paradox'), reduces VEGF expression and angiogenesis in endothelial cells⁹⁸ and tumour cells⁹⁹, and potentially reduces tumour growth and metastasis in tumour-bearing mice⁹⁸. In endothelial cells, resveratrol inhibits VEGF-induced angiogenesis by interrupting Src-dependent VE-cadherin and β -catenin tyrosine phosphorylation, retaining these adherens junction proteins at the cell–cell junctions and preserving endothelial barrier function¹⁰⁰.

Conclusions and future perspectives

Clearly, VEGF-induced permeability did not evolve as a factor to exacerbate the deleterious effects of heart attacks or strokes, so what might its primary physiological function be? VEGF-induced vascular leak at the site of infection could promote extravasation of proteins such as fibrinogen, which polymerizes to form fibrin in the extracellular space and provides an impenetrable barrier to contain an infection or the malignant growth of cells. Indeed, fibrin is typically found surrounding tumours, areas of wound repair, or sites of infection⁷. However, during evolution, it would not have been possible to select against the negative aspects of VEGF-induced vascular leak during cancer or ischaemic injury, because these pathologies usually occur later in life after child-bearing years and are likely to be a function of a Western diet, environmental conditions and/or old age.

Recent clinical efforts are underway to block VEGF-mediated leak pharmacologically in patients after acute myocardial infarction. Such efforts may have a significant impact on reducing tissue injury and thereby minimizing the long-term consequences in this patient population. A similar therapeutic option may become available for stroke patients or individuals with retinal oedema. Blocking VEGF-mediated vascular leak may also have a profound effect in cancer patients by reducing the metastatic entry of tumour cells into or out of the vascular circulation.

1. Senger, D. R. *et al.* Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid. *Science* **219**, 983–985 (1983).
2. Ferrara, N., Gerber, H. P. & LeCouter, J. The biology of VEGF and its receptors. *Nature Med.* **9**, 669–676 (2003).
3. Carmeliet, P. *et al.* Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* **380**, 435–439 (1996).
4. Ferrara, N. *et al.* Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* **380**, 439–442 (1996).
5. Ferrara, N., Hillan, K. J., Gerber, H. P. & Novotny, W. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nature Rev. Drug Discov.* **3**, 391–400 (2004).
6. Eliceiri, B. P. *et al.* Selective requirement for Src kinases during VEGF-induced angiogenesis and vascular permeability. *Mol. Cell* **4**, 915–924 (1999).
7. Brown, L. F., Dvorak, A. M. & Dvorak, H. F. Leaky vessels, fibrin deposition, and fibrosis: a sequence of events common to solid tumors and to many other types of disease. *Am. Rev. Respir. Dis.* **140**, 1104–1107 (1989).
8. van Bruggen, N. *et al.* VEGF antagonism reduces edema formation and tissue damage after ischemia/reperfusion injury in the mouse brain. *J. Clin. Invest.* **104**, 1613–1620 (1999).
9. Weis, S. *et al.* Src blockade stabilizes a Flk/cadherin complex, reducing edema and tissue injury following myocardial infarction. *J. Clin. Invest.* **113**, 885–894 (2004).
10. Paul, R. *et al.* Src deficiency or blockade of Src activity in mice provides cerebral protection following stroke. *Nature Med.* **7**, 222–227 (2001).
11. Jain, R. K. Transport of molecules across tumour vasculature. *Cancer Metastasis Rev.* **6**, 559–593 (1987).
12. Dvorak, H. F. Leaky tumour vessels: consequences for tumour stroma generation and for solid tumour therapy. *Prog. Clin. Biol. Res.* **354A**, 317–330 (1990).
13. Roberts, W. G. & Palade, G. E. Increased microvascular permeability and endothelial fenestration induced by vascular endothelial growth factor. *J. Cell Sci.* **108**, 2369–2379 (1995).
14. Esser, S. *et al.* Vascular endothelial growth factor induces endothelial fenestrations *in vitro*. *J. Cell Biol.* **140**, 947–959 (1998).
15. Feng, D. *et al.* Pathways of macromolecular extravasation across microvascular endothelium in response to VPF/VEGF and other vasoactive mediators. *Microcirculation* **6**, 23–44 (1999).
16. Hobbs, S. K. *et al.* Regulation of transport pathways in tumour vessels: role of tumour type and microenvironment. *Proc. Natl Acad. Sci. USA* **95**, 4607–4612 (1998).

17. Monsky, W. L. *et al.* Augmentation of transvascular transport of macromolecules and nanoparticles in tumors using vascular endothelial growth factor. *Cancer Res.* **59**, 4129–4135 (1999).
18. Weis, S., Cui, J., Barnes, L. & Cheresh, D. Endothelial barrier disruption by VEGF-mediated Src activity potentiates tumour cell extravasation and metastasis. *J. Cell Biol.* **167**, 223–229 (2004).
19. Salgado, R. *et al.* Platelets and vascular endothelial growth factor (VEGF): a morphological and functional study. *Angiogenesis* **4**, 37–43 (2001).
20. Cohen, A. W., Carbajal, J. M. & Schaeffer, R. C. Jr VEGF stimulates tyrosine phosphorylation of β -catenin and small-pore endothelial barrier dysfunction. *Am. J. Physiol.* **277**, H2038–H2049 (1999).
21. Moore, T. M. *et al.* Receptor-dependent activation of store-operated calcium entry increases endothelial cell permeability. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **279**, L691–L698 (2000).
22. Esser, S., Lampugnani, M. G., Corada, M., Dejana, E. & Risau, W. Vascular endothelial growth factor induces VE-cadherin tyrosine phosphorylation in endothelial cells. *J. Cell Sci.* **111**, 1853–1865 (1998).
23. Fischer, S. *et al.* Hypoxia induces permeability in brain microvessel endothelial cells via VEGF and NO. *Am. J. Physiol.* **276**, C812–C820 (1999).
24. Chang, Y. S. *et al.* Effect of vascular endothelial growth factor on cultured endothelial cell monolayer transport properties. *Microvasc. Res.* **59**, 265–277 (2000).
25. Miles, A. A. & Miles, E. M. Vascular reactions to histamine, histamine-liberator and leukotaxine in the skin of guinea pigs. *J. Physiol. (Lond.)* **118**, 228–257 (1952).
26. Liao, F. *et al.* Selective targeting of angiogenic tumour vasculature by vascular endothelial-cadherin antibody inhibits tumour growth without affecting vascular permeability. *Cancer Res.* **62**, 2567–2575 (2002).
27. Bates, D. O., Lodwick, D. & Williams, B. Vascular endothelial growth factor and microvascular permeability. *Microcirculation* **6**, 83–96 (1999).
28. McDonald, D. M., Thurston, G. & Baluk, P. Endothelial gaps as sites for plasma leakage in inflammation. *Microcirculation* **6**, 7–22 (1999).
29. Feng, D., Nagy, J. A., Dvorak, A. M. & Dvorak, H. F. Different pathways of macromolecule extravasation from hyperpermeable tumour vessels. *Microvasc. Res.* **59**, 24–37 (2000).
30. Ishida, A. *et al.* Expression of vascular endothelial growth factor receptors in smooth muscle cells. *J. Cell. Physiol.* **188**, 359–368 (2001).
31. Deckers, M. M. *et al.* Expression of vascular endothelial growth factors and their receptors during osteoblast differentiation. *Endocrinology* **141**, 1667–1674 (2000).
32. Takahashi, N. *et al.* Vascular endothelial growth factor induces activation and subcellular translocation of focal adhesion kinase (p125FAK) in cultured rat cardiac myocytes. *Circ. Res.* **84**, 1194–1202 (1999).
33. Chintalgattu, V., Nair, D. M. & Katwa, L. C. Cardiac myofibroblasts: a novel source of vascular endothelial growth factor (VEGF) and its receptors Flt-1 and KDR. *J. Mol. Cell. Cardiol.* **35**, 277–286 (2003).
34. Carmeliet, P. & Storkbaum, E. Vascular and neuronal effects of VEGF in the nervous system: implications for neurological disorders. *Semin. Cell Dev. Biol.* **13**, 39–53 (2002).
35. Xie, K., Wei, D., Shi, Q. & Huang, S. Constitutive and inducible expression and regulation of vascular endothelial growth factor. *Cytokine Growth Factor Rev.* **15**, 297–324 (2004).
36. Reynolds, L. E. *et al.* Enhanced pathological angiogenesis in mice lacking $\beta 3$ integrin or $\beta 3$ and $\beta 5$ integrins. *Nature Med.* **8**, 27–34 (2002).
37. Friedlander, M. *et al.* Definition of two angiogenic pathways by distinct αv integrins. *Science* **270**, 1500–1502 (1995).
38. Eliceiri, B. P. *et al.* Src-mediated coupling of focal adhesion kinase to integrin $\alpha v \beta 5$ in vascular endothelial growth factor signalling. *J. Cell Biol.* **157**, 149–160 (2002).
39. Robinson, S. D., Reynolds, L. E., Wyder, L., Hicklin, D. J. & Hodivala-Dilke, K. M. $\beta 3$ -Integrin regulates vascular endothelial growth factor-A-dependent permeability. *Thromb. Vasc. Biol.* **24**, 2108–2114 (2004).
40. Borges, E., Jan, Y. & Ruoslahti, E. Platelet-derived growth factor receptor beta and vascular endothelial growth factor receptor 2 bind to the $\beta 3$ integrin through its extracellular domain. *J. Biol. Chem.* **275**, 39867–39873 (2000).
41. Reynolds, A. R. *et al.* Elevated Flk1 (vascular endothelial growth factor receptor 2) signalling mediates enhanced angiogenesis in $\beta 3$ -integrin-deficient mice. *Cancer Res.* **64**, 8643–8650 (2004).
42. Dejana, E., Corada, M. & Lampugnani, M. G. Endothelial cell-to-cell junctions. *FASEB J.* **9**, 910–918 (1995).
43. Suarez, S. & Ballmer-Hofer, K. VEGF transiently disrupts gap junctional communication in endothelial cells. *J. Cell Sci.* **114**, 1229–1235 (2001).
44. Antonetti, D. A., Barber, A. J., Hollinger, L. A., Wolpert, E. B. & Gardner, T. W. Vascular endothelial growth factor induces rapid phosphorylation of tight junction proteins occludin and zonula occluden 1. A potential mechanism for vascular permeability in diabetic retinopathy and tumors. *J. Biol. Chem.* **274**, 23463–23467 (1999).
45. Pedram, A., Razandi, M. & Levin, E. R. Deciphering vascular endothelial cell growth factor/vascular permeability factor signalling to vascular permeability. Inhibition by atrial natriuretic peptide. *J. Biol. Chem.* **277**, 44385–44398 (2002).
46. Corada, M. *et al.* Vascular endothelial-cadherin is an important determinant of microvascular integrity *in vivo*. *Proc. Natl Acad. Sci. USA* **96**, 9815–9820 (1999).
47. Kevil, C. G., Okayama, N. & Alexander, J. S. H_2O_2 -mediated permeability II: importance of tyrosine phosphatase and kinase activity. *Am. J. Physiol. Cell Physiol.* **281**, C1940–C1947 (2001).
48. Nwariaku, F. E. *et al.* Tyrosine phosphorylation of vascular endothelial cadherin and the regulation of microvascular permeability. *Surgery* **132**, 180–185 (2002).
49. Lambeng, N. *et al.* Vascular endothelial-cadherin tyrosine phosphorylation in angiogenic and quiescent adult tissues. *Circ. Res.* **96**, 384–391 (2005).
50. Suyama, K., Shapiro, I., Guttman, M. & Hazan, R. B. A signalling pathway leading to metastasis is controlled by N-cadherin and the FGF receptor. *Cancer Cell* **2**, 301–314 (2002).
51. Fedor-Chaikin, M., Hein, P. W., Stewart, J. C., Brackenbury, R. & Kinch, M. S. E-cadherin binding modulates EGF receptor activation. *Cell Commun. Adhes.* **10**, 105–118 (2003).
52. Avizienyte, E., Fincham, V. J., Brunton, V. G. & Frame, M. C. Src SH3/2 domain-mediated peripheral accumulation of Src and phospho-myosin is linked to de-regulation of E-cadherin and the epithelial-mesenchymal transition. *Mol. Biol. Cell* **15**, 2794–2803 (2004).
53. Roura, S., Miravet, S., Piedra, J., Garcia de Herreros, A. & Dunach, M. Regulation of E-cadherin/Catenin association by tyrosine phosphorylation. *J. Biol. Chem.* **274**, 36734–36740 (1999).
54. Ozawa, M. & Ohkubo, T. Tyrosine phosphorylation of p120(ctn) in v-Src transfected L cells depends on its association with E-cadherin and reduces adhesion activity. *J. Cell Sci.* **114**, 503–512 (2001).
55. Xu, Y. & Carpenter, G. Identification of cadherin tyrosine residues that are phosphorylated and mediate Shc association. *J. Cell. Biochem.* **75**, 264–271 (1999).
56. Balsamo, J. *et al.* Regulated binding of PTP1B-like phosphatase to N-cadherin: control of cadherin-mediated adhesion by dephosphorylation of β -catenin. *J. Cell Biol.* **134**, 801–813 (1996).
57. Brady-Kalnay, S. M. *et al.* Dynamic interaction of PTP μ with multiple cadherins *in vivo*. *J. Cell Biol.* **141**, 287–296 (1998).
58. Nawroth, R. *et al.* VE-PTP and VE-cadherin ectodomains interact to facilitate regulation of phosphorylation and cell contacts. *EMBO J.* **21**, 4885–4895 (2002).
59. Grazia Lampugnani, M. *et al.* Contact inhibition of VEGF-induced proliferation requires vascular endothelial cadherin, β -catenin, and the phosphatase DEP-1/CD148. *J. Cell Biol.* **161**, 793–804 (2003).
60. Ukropec, J. A., Hollinger, M. K., Salva, S. M. & Woolkalis, M. J. SHP2 association with VE-cadherin complexes in human endothelial cells is regulated by thrombin. *J. Biol. Chem.* **275**, 5983–5986 (2000).
61. Kroll, J. & Waltenberger, J. The vascular endothelial growth factor receptor KDR activates multiple signal transduction pathways in porcine aortic endothelial cells. *J. Biol. Chem.* **272**, 32521–32527 (1997).
62. Wu, H. M., Yuan, Y., Zawieja, D. C., Tinsley, J. & Granger, H. J. Role of phospholipase C, protein kinase C, and calcium in VEGF-induced venular hyperpermeability. *Am. J. Physiol.* **276**, H535–H542 (1999).
63. Behzadian, M. A. *et al.* VEGF-induced paracellular permeability in cultured endothelial cells involves urokinase and its receptor. *FASEB J.* **17**, 752–754 (2003).
64. Langille, B. L. Morphologic responses of endothelium to shear stress: reorganization of the adherens junction. *Microcirculation* **8**, 195–206 (2001).
65. Chen, K. D. *et al.* Mechanotransduction in response to shear stress. Roles of receptor tyrosine kinases, integrins, and Shc. *J. Biol. Chem.* **274**, 18393–18400 (1999).
66. Stockton, R. A., Schaefer, E. & Schwartz, M. A. p21-activated kinase regulates endothelial permeability through modulation of contractility. *J. Biol. Chem.* **279**, 46621–46630 (2004).
67. Wu, M. H., Guo, M., Yuan, S. Y. & Granger, H. J. Focal adhesion kinase mediates porcine venular hyperpermeability elicited by vascular endothelial growth factor. *J. Physiol. (Lond.)* **552**, 691–699 (2003).
68. Yue, X. & Tomanek, R. J. Stimulation of coronary vasculogenesis/angiogenesis by hypoxia in cultured embryonic hearts. *Dev. Dyn.* **216**, 28–36 (1999).
69. Marti, H. J. *et al.* Hypoxia-induced vascular endothelial growth factor expression precedes neovascularization after cerebral ischemia. *Am. J. Pathol.* **156**, 965–976 (2000).
70. Li, J. *et al.* VEGF, flk-1, and flt-1 expression in a rat myocardial infarction model of angiogenesis. *Am. J. Physiol.* **270**, H1803–H1811 (1996).
71. Fukuda, S. *et al.* Angiogenic signal triggered by ischemic stress induces myocardial repair in rat during chronic infarction. *J. Mol. Cell. Cardiol.* **36**, 547–559 (2004).
72. Sasaki, H. *et al.* Hypoxic preconditioning triggers myocardial angiogenesis: a novel approach to enhance contractile functional reserve in rat with myocardial infarction. *J. Mol. Cell. Cardiol.* **34**, 335–348 (2002).
73. Shyu, K. G. *et al.* Intramyocardial injection of naked DNA encoding HIF-1 α /VP16 hybrid to enhance angiogenesis in an acute myocardial infarction model in the rat. *Cardiovasc. Res.* **54**, 576–583 (2002).
74. Kawamoto, A. *et al.* Intramyocardial transplantation of autologous endothelial progenitor cells for therapeutic neovascularization of myocardial ischemia. *Circulation* **107**, 461–468 (2003).
75. Kawamoto, A. *et al.* Synergistic effect of bone marrow mobilization and

- vascular endothelial growth factor-2 gene therapy in myocardial ischemia. *Circulation* **110**, 1398–1405 (2004).
76. Lee, R. J. *et al.* VEGF gene delivery to myocardium: deleterious effects of unregulated expression. *Circulation* **102**, 898–901 (2000).
 77. Tong, R. T. *et al.* Vascular normalization by vascular endothelial growth factor receptor 2 blockade induces a pressure gradient across the vasculature and improves drug penetration in tumors. *Cancer Res.* **64**, 3731–3736 (2004).
 78. Willett, C. G. *et al.* Direct evidence that the VEGF-specific antibody bevacizumab has antivasculature effects in human rectal cancer. *Nature Med.* **10**, 145–147 (2004).
 79. Kuroi, K. & Toi, M. Circulating angiogenesis regulators in cancer patients. *Int. J. Biol. Markers* **16**, 5–26 (2001).
 80. Hormbrey, E. *et al.* A critical review of vascular endothelial growth factor (VEGF) analysis in peripheral blood: is the current literature meaningful? *Clin. Exp. Metastasis* **19**, 651–663 (2002).
 81. Yuan, F. *et al.* Time-dependent vascular regression and permeability changes in established human tumour xenografts induced by an anti-vascular endothelial growth factor/vascular permeability factor antibody. *Proc. Natl Acad. Sci. USA* **93**, 14765–14770 (1996).
 82. Hurwitz, H. *et al.* Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N. Engl. J. Med.* **350**, 2335–2342 (2004).
 83. Winkler, F. *et al.* Kinetics of vascular normalization by VEGFR2 blockade governs brain tumour response to radiation: role of oxygenation, angiopoietin-1, and matrix metalloproteinases. *Cancer Cell* **6**, 553–563 (2004).
 84. Jain, R. K. Normalization of tumour vasculature: An emerging concept in antiangiogenic therapy. *Science* **307**, 58–62 (2005).
 85. Mukhopadhyay, D., Tsiokas, L. & Sukhatme, V. P. Wild-type p53 and v-Src exert opposing influences on human vascular endothelial growth factor gene expression. *Cancer Res.* **55**, 6161–6165 (1995).
 86. Mukhopadhyay, D. & Datta, K. Multiple regulatory pathways of vascular permeability factor/vascular endothelial growth factor (VPF/VEGF) expression in tumors. *Semin. Cancer Biol.* **14**, 123–130 (2004).
 87. Izumi, Y., Xu, L., di Tomaso, E., Fukumura, D. & Jain, R. K. Tumour biology: herceptin acts as an anti-angiogenic cocktail. *Nature* **416**, 279–280 (2002).
 88. Baish, J. W., Netti, P. A. & Jain, R. K. Transmural coupling of fluid flow in microcirculatory network and interstitium in tumors. *Microvasc. Res.* **53**, 128–141 (1997).
 89. Tsuzuki, Y. *et al.* Vascular endothelial growth factor (VEGF) modulation by targeting hypoxia-inducible factor-1 α $\rightarrow$ hypoxia response element $\rightarrow$ VEGF cascade differentially regulates vascular response and growth rate in tumors. *Cancer Res.* **60**, 6248–6252 (2000).
 90. Akiyama, C. *et al.* Src family kinase inhibitor PP1 improves motor function by reducing edema after spinal cord contusion in rats. *Acta Neurochir. Suppl.* **86**, 421–423 (2003).
 91. Mura, M., dos Santos, C. C., Stewart, D. & Liu, M. Vascular endothelial growth factor and related molecules in acute lung injury. *J. Appl. Physiol.* **97**, 1605–1617 (2004).
 92. Vaquero, J., Chung, C. & Blei, A. T. Brain edema in acute liver failure. A window to the pathogenesis of hepatic encephalopathy. *Ann. Hepatol.* **2**, 12–22 (2003).
 93. Caldwell, R. B. *et al.* Vascular endothelial growth factor and diabetic retinopathy: pathophysiological mechanisms and treatment perspectives. *Diabetes Metab. Res. Rev.* **19**, 442–455 (2003).
 94. Conklin, B. S., Zhao, W., Zhong, D. S. & Chen, C. Nicotine and cotinine up-regulate vascular endothelial growth factor expression in endothelial cells. *Am. J. Pathol.* **160**, 413–418 (2002).
 95. Zhu, B. Q. *et al.* Second hand smoke stimulates tumour angiogenesis and growth. *Cancer Cell* **4**, 191–196 (2003).
 96. Liao, H. F. *et al.* Inhibitory effect of caffeic acid phenethyl ester on angiogenesis, tumour invasion, and metastasis. *J. Agric. Food Chem.* **51**, 7907–7912 (2003).
 97. Tang, F. Y., Nguyen, N. & Meydani, M. Green tea catechins inhibit VEGF-induced angiogenesis *in vitro* through suppression of VE-cadherin phosphorylation and inactivation of Akt molecule. *Int. J. Cancer* **106**, 871–878 (2003).
 98. Kimura, Y. & Okuda, H. Resveratrol isolated from *Polygonum cuspidatum* root prevents tumour growth and metastasis to lung and tumour-induced neovascularization in Lewis lung carcinoma-bearing mice. *J. Nutr.* **131**, 1844–1849 (2001).
 99. Cao, Z., Fang, J., Xia, C., Shi, X. & Jiang, B. H. trans-3,4,5'-Trihydroxystibene inhibits hypoxia-inducible factor 1 α and vascular endothelial growth factor expression in human ovarian cancer cells. *Clin. Cancer Res.* **10**, 5253–5263 (2004).
 100. Lin, M. T., Yen, M. L., Lin, C. Y. & Kuo, M. L. Inhibition of vascular endothelial growth factor-induced angiogenesis by resveratrol through interruption of Src-dependent vascular endothelial cadherin tyrosine phosphorylation. *Mol. Pharmacol.* **64**, 1029–1036 (2003).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to D.A.C. (dcheresh@ucsd.edu).

Structures of complement component C3 provide insights into the function and evolution of immunity

Bert J. C. Janssen¹, Eric G. Huizinga¹, Hans C. A. Raaijmakers¹, Anja Roos², Mohamed R. Daha², Kristina Nilsson-Ekdahl^{3,4}, Bo Nilsson³ & Piet Gros¹

The mammalian complement system is a phylogenetically ancient cascade system that has a major role in innate and adaptive immunity. Activation of component C3 (1,641 residues) is central to the three complement pathways and results in inflammation and elimination of self and non-self targets. Here we present crystal structures of native C3 and its final major proteolytic fragment C3c. The structures reveal thirteen domains, nine of which were unpredicted, and suggest that the proteins of the α 2-macroglobulin family evolved from a core of eight homologous domains. A double mechanism prevents hydrolysis of the thioester group, essential for covalent attachment of activated C3 to target surfaces. Marked conformational changes in the α -chain, including movement of a critical interaction site through a ring formed by the domains of the β -chain, indicate an unprecedented, conformation-dependent mechanism of activation, regulation and biological function of C3.

The mammalian complement system mediates inflammation by generating anaphylatoxins to elicit chemotaxis and cell activation, and by promoting phagocytosis, degranulation and cell lysis. Its main functions include host defence against microorganisms, elimination of immune complexes and apoptotic cells, and facilitating adaptive immune responses^{1,2}. The pivotal complement factor C3 is the convergence point for the classical, lectin and alternative pathways of complement activation^{1,2}.

C3 (187 kDa) emerged over 700 million years ago³ and belongs to the α 2-macroglobulin (α 2M) family. These large (approximately 1,400–1,800 residue) proteins are characterized by homologous sequence features, including a unique thioester motif and a central, highly variable part that is functionally important⁴. Its members, such as the complement factors C3, C4 and C5, the proteinase inhibitor α 2M and the insect and nematode thioester-containing proteins (TEPs), have important roles in the immune response in metazoans, considerably pre-dating the emergence of immunoglobulins⁵.

C3 interacts with a large number of complement factors (for example, proteases, receptors and regulators) and non-complement proteins (for example, viral and bacterial proteins) via distinct binding sites. The function of C3 is regulated by conformational changes induced by sequential proteolytic cleavages (Fig. 1). Cleavage of mature C3 is mediated by enzyme complexes (that is, convertases), and generates the anaphylatoxin C3a (9 kDa) and the major fragment C3b (177 kDa)⁶ (Fig. 1c). This step exposes a hidden thioester⁷ and multiple cryptic binding sites in C3b for interacting complement proteins⁸. Nascent C3b is able to bind covalently to cell and other target surfaces via the exposed thioester⁹. Amplification of complement activity is achieved by association of surface-bound C3b and pro-enzyme factor B¹⁰, yielding the short-lived ($t_{1/2} \approx 90$ s)¹¹ C3 convertase C3bBb of the alternative pathway. Additional cleavages in the α -chain of C3b (generating iC3b) mediated by factor I in

association with soluble^{12,13} or membrane-bound^{13,14} co-factors prevent further convertase formation and profoundly alter the function of the protein⁸. The first two cleavages release C3f (2 kDa)¹⁵ and the third cleavage in the remaining iC3b liberates C3c (135 kDa) from the target-bound C3dg (40 kDa) fragment¹⁶.

Structure determination of C3c and C3

We determined the structures of human, native C3 (3.3 Å resolution) and its main proteolytic fragment C3c (2.4 Å resolution), which comprises 72% of C3. First, we determined the structure of C3c purified from outdated human plasma. The structure was solved by single-isomorphous replacement with anomalous scattering (SIRAS) and two-wavelength anomalous dispersion (MAD) phasing but required multi-crystal averaging¹⁷ with nine partial masks to obtain an interpretable map. Second, we determined the structure of full-length C3 purified from fresh human plasma. We solved this structure by molecular replacement. Initial positioning of C3c, or any of its fragments, was unsuccessful. However, the α 6– α 6 helical structure of C3d¹⁸ (18% of C3, in a protein otherwise rich in β -strands) was positioned correctly by Phaser¹⁹. Subsequently, domains of C3c were added gradually, yielding 85% of the model, which was completed by model building. The final refined models had R and R_{free} values of 22.3 and 27.9% (C3c) and 23.6 and 29.5% (C3); see Methods and Table 1; see also Supplementary Fig. 1 and Supplementary Table 1.

Domain organization

C3 consists of two chains— β (residues 1–645) and α (residues 650–1641) (ref. 20)—that together form 13 domains, whereas C3c consists of three chains (the β -chain and two fragments of the α -chain) that form ten domains (Fig. 1a–c; see also Supplementary Fig. 2). Surprisingly, one domain is formed by parts of both the β - and

¹Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Faculty of Science, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

²Department of Nephrology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. ³Department of Clinical Immunology, University Hospital, SE-751 85 Uppsala, Sweden. ⁴Department of Chemistry and Biomedical Sciences, University of Kalmar, SE-391 82 Kalmar, Sweden.

α -chains. In addition, each chain forms six domains by itself. Eight domains (5.5 of β and 2.5 of α) exhibit a fibronectin-type-3-like core fold (Supplementary Fig. 2d, e); however, no sequence homology is apparent among these domains. In analogy to immunoglobulin domains, we refer to them as macroglobulin (MG) domains. Residues 1–534 form five MG domains, MG1–MG5 (Supplementary Fig. 3). Residues 535–577 form one half, denoted MG6 $^{\beta}$, of the β/α intertwined MG6 domain. Together the chain of six MG domains forms 1.5 turns of helical coil as in a key ring (see domains of β -chain in Fig. 1c). Residues 578–645 of the β -chain exit from MG6. They loop downwards and through the ring, forming three helices in an extended configuration and one β -strand that aligns with the first strand of MG1. Three aromatic residues from the third helix and the β -strand of this linker domain, called LNK, form a small hydrophobic core. This structural element is wedged in between domains MG1, MG4 and MG5. The α -chain starts with the anaphylatoxin

(ANA) domain (residues 650–726), which when cleaved off forms the C3a fragment. The structure of this domain is similar to the crystal structure of C3a (ref. 21), although the amino-terminal α -helical region was not resolved in that structure (Supplementary Fig. 3h). The subsequent scissile bond Arg 726–Ser 727 sits in a surface-exposed, disordered loop (residues 720–729). Following this, there is an extended loop (residues 730–745) that connects ANA to MG6. Loop 727–745 forms the N-terminal region of the cleaved α -chain, denoted α' , in C3b; we refer to this loop as α' NT. Residues 746–806 form MG6 $^{\alpha}$, which complements MG6 $^{\beta}$ in the formation of MG6. MG6 $^{\alpha}$ is followed by a seventh MG domain, MG7 (residues 807–911). The polypeptide chain that is excised by factor I to yield fragments C3dg and C3f is located between MG7 and an eighth MG domain, MG8 (residues 1331–1474). MG8 corresponds to the receptor-binding domain in $\alpha 2M$; as expected, the structure of MG8 is homologous to this domain²² (Supplementary Fig. 3p).

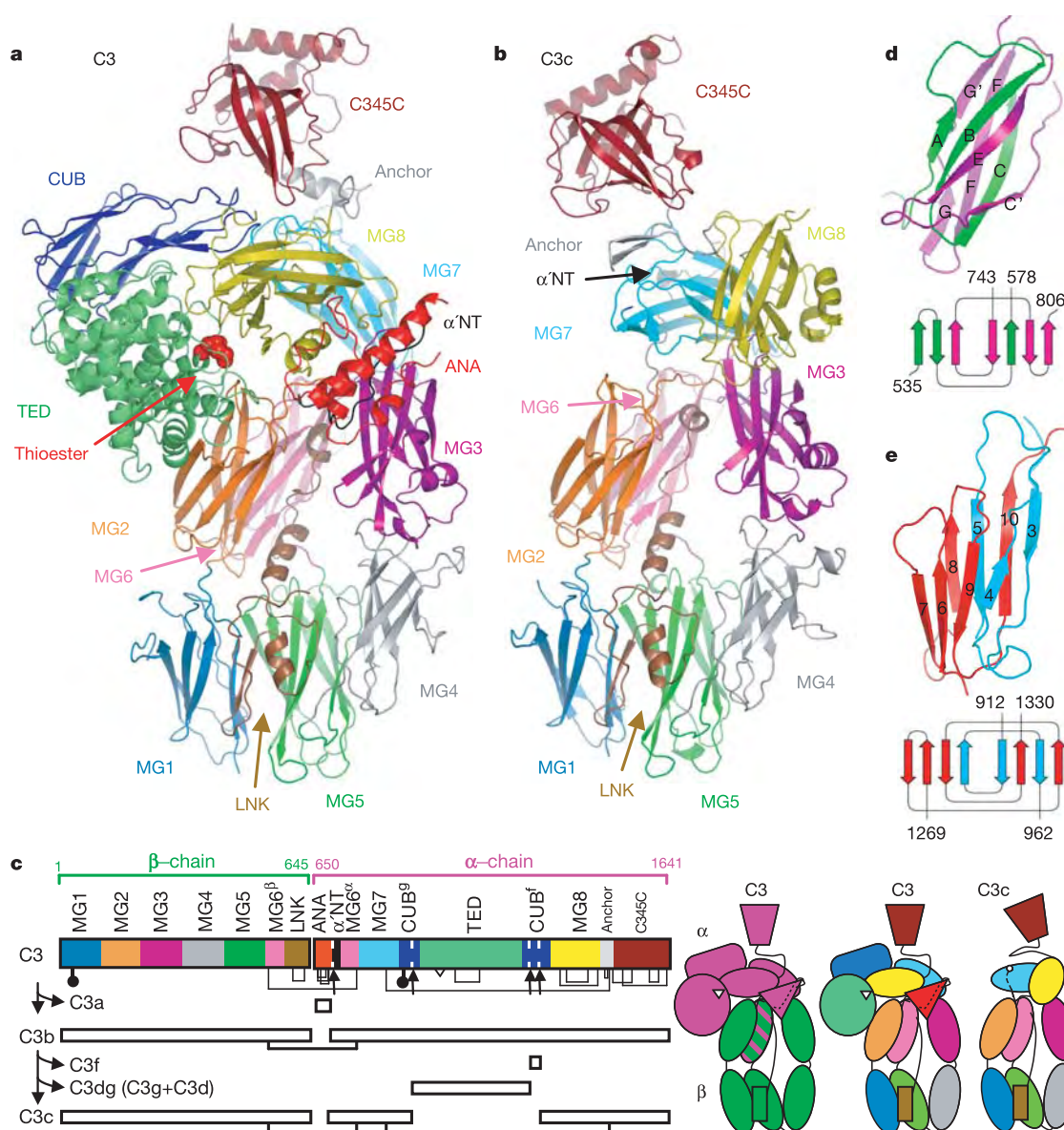


Figure 1 | Structures of human complement components C3 and C3c.

a, b, Ribbon representation of native C3 (13 domains) and C3c (10 domains), respectively. Also shown are intact thioester (red spheres), anchor region (grey) and α' NT (black). **c**, Domain sequence and arrangements in C3 and C3c. The colour scheme matches that in **a, b**. Shown are the thioester site (white triangle), disulphide bridges, glycan positions (for details see

Supplementary Figs 3 and 6) and cleavage sites. Sequential proteolysis from C3 to C3c is indicated. **d, e**, Intertwined domains. **d**, MG6 intertwines the β - and α -chain of mature C3. MG6 $^{\beta}$, green; MG6 $^{\alpha}$, pink. **e**, Intertwined CUB. CUB $^{\beta}$, cyan; CUB $^{\alpha}$, red. Fibronectin-type 3 and CUB strand numbering are indicated.

Table 1 | Refinement statistics

	C3c	C3
Resolution (Å)	40–2.4	40–3.3
$R_{\text{work}}/R_{\text{free}}$ (%)	22.3/27.9	23.6/29.5
Number of atoms		
Protein	17,686	12,859
Ligand/ion	90	–
Water	612	–
B factors (Å ²)		
Protein	46	93
Ligand/ion	67	–
Water	43	–
Root mean square deviations		
Bond lengths (Å)	0.012	0.002
Bond angles (degrees)	1.45	0.532

The cleavage sites yielding C3g, C3d and C3f^{15,16} do not coincide fully with domain boundaries. Residues 912–962 and 1269–1330, 63% of C3g and all of C3f, together form one domain that displays a CUB fold; we refer to the two separate parts as CUB^g and CUB^f, respectively (Fig. 1e). The remaining part of C3g and C3d (residues 963–1268) together form a thioester-containing domain (TED). Despite a major rearrangement of the N-terminal region, the $\alpha 6$ – $\alpha 6$ fold is conserved between TED and C3d¹⁸ (Supplementary Fig. 3m). Finally, residues 1496–1641 form a carboxy-terminal C345C domain with a netrin-like fold. This domain is covalently linked to MG8 by the polypeptide chain and to MG7 by a disulphide bond (Cys 851–Cys 1491)²³ in what we refer to as an anchor region (residues 1475–1495). The overall structure of C345C is similar to the recent NMR-solution structure of the C345C domain of C5 (ref. 24; see also Supplementary Fig. 3r). In total, nine out of thirteen domains were unpredicted (MG1–MG7, LNK and CUB). All domains together form an irregularly shaped C3 and a disc-shaped C3c that lacks the ANA, CUB and TED domains, with strong dipolar surface-charge distributions (Supplementary Fig. 4).

The α -chain undergoes major domain rearrangements

There are marked conformational differences between C3 and C3c. The overall shape of the β -ring is well conserved; the only large change in MG1–MG6 and LNK is a 15° rotation and a 4.3 Å (centre of mass) shift of MG3 (Fig. 2; see also Supplementary Table 2). The domains of the α -chain, however, undergo large rearrangements when the molecule is processed from C3 into C3c. MG7 and MG8

almost swap places in the overall structure: they rotate and translate by 35° and 7.8 Å, and 62° and 24 Å, respectively. The C345C swivels 32° and moves 10.1 Å, resulting in a torque motion acting on the anchor region. The anchor region, with its internal Cys 1484–Cys 1489 disulphide bond, alters its conformation completely, from an α -helix in C3 to a β -hairpin in C3c (Fig. 2b; see also Supplementary Figs 2c and 3q). Overall, the structural differences between C3 and C3c suggest that the β -ring forms a relatively stable molecular platform for the structurally adaptable α -chain to undergo induced conformational changes.

Implications for evolutionary events

Intertwining of distant parts of the amino-acid sequence owing to inclusion of a domain into a loop is indicative of a gene insertion event. This situation occurs twice in C3 (Fig. 1d, e). MG6 ^{β/α} has an insert of 165 residues in loop β C– β C' and CUB ^{β/α} has an insert of 307 residues in loop β 5– β 6. The MG6 insert forms the linker domain, the processing site 646–RRRR–649 (present in proC3), the anaphylatoxin domain, scissile bond Arg 726–Ser 727 and the α NT, which contains four acidic residues important for factor B binding to C3b²⁵. Because the MG6 insert includes the proC3 processing site, it follows that MG6 is formed by two separate chains in mature C3. The insert into the CUB domain is the TED domain with the critical thioester group. Moreover, we hypothesize that the CUB domain itself is a putative insert into the linker between MG7 and MG8 (and hence the obvious signal of a domain formed by distant parts of the sequence is lacking). Notably, critical C3 elements such as the thioester, the anaphylatoxin and the proteolytic cleavage sites lie within these putative inserts.

Comparison of C3 with other members of the $\alpha 2$ M family (Fig. 3) shows that the MG6 insert encompasses the central variable region, which in $\alpha 2$ M contains the bait region²⁶. In phylogenetically distant (insect and nematode) TEPs this sequence is hypervariable²⁷. In contrast, all sequences contain CUB^g-, TED- and CUB^f-like elements, the hallmark of $\alpha 2$ M family members. Homologous sequences lacking these elements are not known. The occurrence of $\alpha 6$ – $\alpha 6$ folds in enzymes¹⁸ indicates perhaps that the TED fold existed separately before it was inserted and became a critical component in the proteins of the $\alpha 2$ M family. The combination of two putative gene insertions and a gene extension of the C-terminal C345C domain²⁸ implies the existence of a hypothetical ancestral molecule comprised of eight MG domains, which possibly arose by gene duplication events. The sequences of the MG domains have diverged to an extent whereby there is no longer any detectable sequence

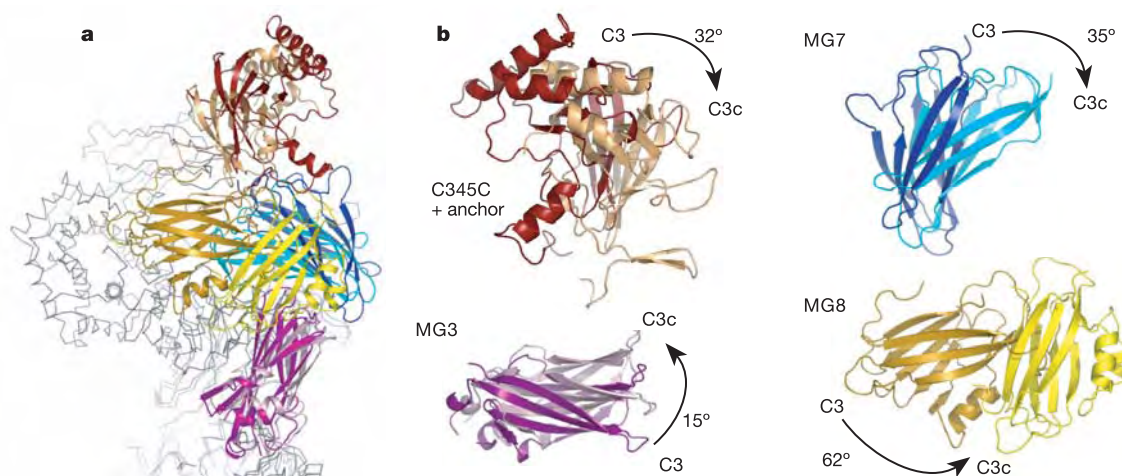


Figure 2 | Differences in domain arrangements between C3 and C3c. **a**, C3 and C3c superposed on the basis of MG1, MG2 and MG4–MG6 of the β -ring shown as C α trace of C3 (grey) and C3c (light grey). Domains that undergo large rearrangements (MG3, MG7, MG8 and C345C) are shown in ribbon representation. The colour scheme matches that of Fig. 1, with dark colours

for C3 and lighter colours for C3c. **b**, Rearrangements observed for domains MG3, MG7, MG8 and C345C; orientations are different compared with **a**, for clarity. MG8 and the anchor region of C345C differ in secondary structure between C3 and C3c (Supplementary Figs 2a and 3p, q).

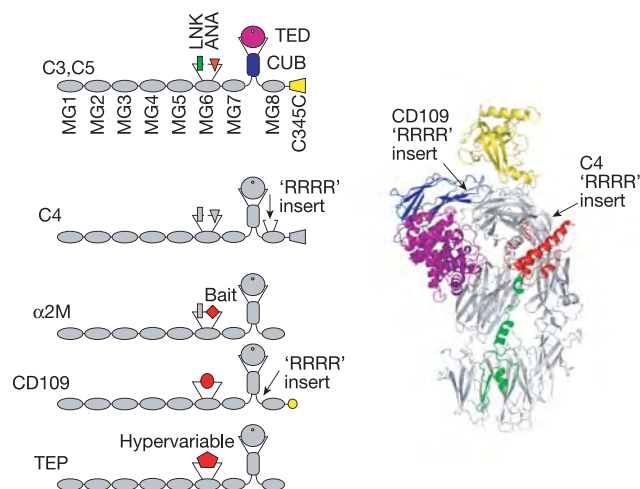


Figure 3 | Domain organization of the $\alpha 2M$ family deduced from the C3 structure. Members of the $\alpha 2M$ family are aligned schematically by domains with LNK, ANA, CUB and TED presented as inserts that potentially arose by gene-insertion events. Mature C4 has an additional tetra-arginine (RRRR) processing site that occurs in an insert of 46 residues in loop $\beta A-\beta B$ of MG8. CD109 (ref. 50) has a tetra-arginine site in the linker between CUB and MG8. Proteins of the $\alpha 2M$ family differ in their composition at the C terminus. $\alpha 2M$ and TEPs lack the C345C domain, whereas CD109 has a GPI anchor here.

homology. Insertion of CUB^g/TED/CUB^f marks the emergence of the $\alpha 2M$ family in host protection, and additional insertion of LNK/RRRR/ANA/ α' NT and extension with C345C marks the emergence of the complement system in the innate immune defence more than 700 million years ago²⁹.

A double mechanism protects the thioester

The highly reactive thioester (formed by the side chains of Cys 988 and Gln 991)²³ is intact in the structure of C3. It is shielded from reacting with water ($t_{1/2} > 6$ days)³⁰ or other small nucleophiles^{31,32} by a hydrophobic/aromatic pocket formed by residues Met 1378,

Tyr 1425 and Tyr 1460 from MG8 and Phe 1047 from the TED (Fig. 4a). These residues are conserved in the $\alpha 2M$ family, except for C5, which lacks the thioester moiety. Although fully buried, the thioester and its protective pocket are positioned at the TED–MG8 interface close to the protein surface.

Upon activation by the proteolytic cleavage of C3 into C3b, the protein reacts more readily with hydroxyl groups than amino groups³³. The altered reactivity is due to a transformation of the thioester into a free thiolate anion, on Cys 988, and formation of an acyl-imidazole intermediate^{34,35} by Gln 991 and His 1104 that is possibly stabilized by Glu 1106 (refs 18, 36). In the structure of C3 His 1104 and Glu 1106 (situated next to a disordered loop) are far from each other and far away from Gln 991 (11.7 Å distance between Gln 991 C δ and His 1104 N ϵ ; Fig. 4b), and hence the catalytic site for reaction with hydroxyls is not present. In the structure of C3d¹⁸, expressed in *Escherichia coli* with a Cys988Ala mutation, His 1104 is only 4.1 Å from Gln 991 in the modelled thioester, and the side chains of His 1104 and Glu 1106 form a hydrogen bond, as would be the case when Glu 1106 stabilizes the acyl-imidazole intermediate (Fig. 4b). In native C3, movement of His 1104 and Glu 1106 towards the thioester is prevented by the MG8–TED interface, thereby blocking the formation of the free thiolate and acyl-imidazole intermediate. This provides a second, specific protection mechanism of the native protein against hydrolysis by water and explains why the reactive moiety in C3 is more resistant to water and less resistant to reaction with small amino nucleophiles^{31,32}.

Activation of C3

The orientation of the TED domain with the thioester pointing inwards seems to be essential for maintaining the protective TED–MG8 interactions in native C3. TED is buttressed in this position by interactions with MG2, MG8 and CUB (Fig. 4c). Cleavage of C3 at Ser 726–Arg 727 removes the ANA domain (which becomes the anaphylatoxin C3a) and yields an activated C3b with an exposed and reactive thioester. Therefore, the ANA domain has a critical role in protecting the thioester in native C3. However, in the structure of C3 no direct contacts are observed between the ANA and TED domains. Although loops of ANA and TED come within ~ 10 Å, there are no charged residues that could give long-range electrostatic

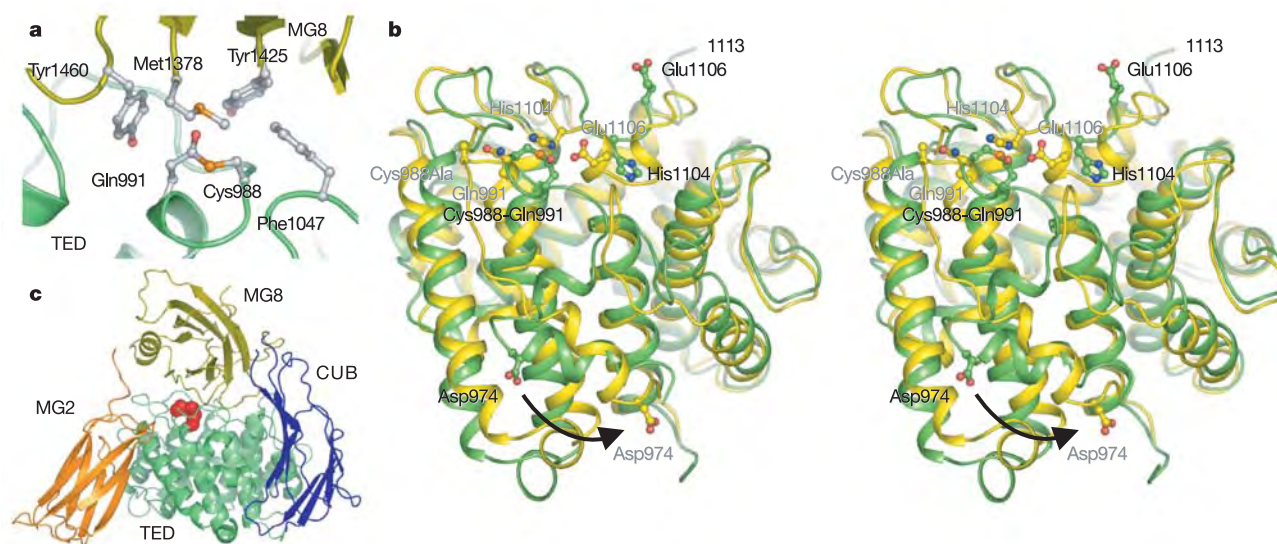


Figure 4 | Interactions of the thioester and the TED domain. **a**, The first thioester protection mechanism involves a hydrophobic/aromatic pocket formed by residues of TED and MG8. Shown are protective residues and thioester (Cys 988–Gln 991) (ball-and-stick representation) with TED (green ribbon) and MG8 (yellow ribbon). **b**, The second thioester protection mechanism blocks reaction with hydroxyls by placing His 1104 and Glu 1106

far from the thioester. Stereo diagram of superposed TED domain (green) and C3d¹⁸ (yellow). Cys 988 was mutated to Ala in C3d¹⁸. Movements in the N-terminal $\alpha 0-\alpha 1$ region between TED and C3d are shown by 15 Å displacement of Asp 974. **c**, Domain–domain interactions stabilize the MG2–TED–CUB–MG8 arrangement. Domain colours match those used in Fig. 1.

attractions. Therefore ANA must protect the thioester in native C3 indirectly. ANA has extensive interactions with MG8. At the other side ANA interacts with MG3 (Fig. 5a) and thus bridges interactions between MG8 and MG3 of the stable β -ring. The ANA domain may serve to keep MG8 in a correct position for interaction with TED and, possibly, to induce a conformation of MG8 that enhances interactions with TED.

MG8 undergoes a conformational and a huge positional change from C3 to C3c (Fig. 2). The domain makes a dramatic swing from the interior of the C3 molecule to the exterior in C3c. This swing moves helix α 1 and surrounding β -strands by 40–50 Å. In C3 these structural elements are hidden and interact with the ANA and TED domains, whereas in C3c they are fully exposed and have changed conformation into a β -strand and two α -helices. This large positional change corresponds well with what is expected for the equivalent receptor-binding domain in α 2M, which alters from hidden to exposed upon encapsulating a proteinase in an α 2M multimer³⁷. In the exposed position its β - α - β motif is thought to bind the α 2M receptor²². In native C3 the β - α - β configuration in MG8 is probably important; it contributes to interaction with TED (notably, residues Lys 1409 and the highly conserved Glu 1411). Furthermore, this MG8 region harbours a binding site for complement protein properdin³⁸. The occluded position in C3 explains why properdin does not bind to native C3. Because properdin stabilizes the C3bBb convertase complex, it follows that MG8 exposes this binding site, possibly in a β - α - α configuration and a swung out position as in C3c.

The observed domain arrangement supports a simple model for exposing the thioester in the activated state of C3. The removal of ANA, yielding C3b, weakens the interactions between MG8 and TED, thereby allowing TED to swing out of its nestled position. Similar, although much slower, conformational changes occur after spontaneous hydrolysis of the thioester in C3, yielding C3(H₂O) (refs 31, 32). In this case, the conformational changes are retarded by the presence of the anaphylatoxin domain. Rotation of TED around its N- and C-terminal connections to CUB exposes TED with the intact thioester in C3b projected outwards. This simple rotation is supported by recent hydrogen–deuterium exchange data on C3 and

C3(H₂O), which indicate that segments 956–968 and 1027–1036 buried in C3 become exposed in C3(H₂O) (ref. 39). Reduced exchange rates for segment 1108–1123 correlate with the disorder–order transition in loop 1107–1112 of C3 versus C3d. This loop lies next to the critical His 1104 and Glu 1106 residues. Possibly, the rotation of TED is linked to the formation of the free thiolate and acyl-imidazole intermediate that requires ~15 Å movements in the N-terminal α 0– α 1 region to establish an active conformation for attachment to a target surface (Fig. 4b).

Implications for convertase formation and regulation

Cleavage of C3 at 726–727 yields a novel N-terminal (that is, α' NT) region in C3b. Four acidic residues (Asp 730, Glu 731, Glu 736 and Glu 737) of the α' NT are important in the formation of the C3 convertase²⁵. In addition, the extended segment 727–767 (that is, α' NT plus the beginning of MG6 α) is important for binding of several regulators^{25,40,41}. In native C3, acidic residues Asp 730, Glu 731, Glu 736 and Glu 737 are shielded by ANA (Fig. 5a), which may explain why factor B cannot bind to native C3. Most surprisingly, the α' NT region (residues 727–744) resides on opposite sides in C3 and C3c (Fig. 5b). From C3 to C3c the α' NT has moved from its position near ANA, through the β -ring and has emerged on the other side in C3c where it lies on the surface of MG7. Sequence conservation of the hinge 745–FPES-748 that connects α' NT to MG6, and of the bridging loop 205–YVLP-208 of the β -ring (Supplementary Fig. 5), suggests that this remarkable rearrangement may be part of a general mechanism of α 2M protein family members.

The striking differences between C3 and C3c concerning the position of the α' NT region and the arrangement of the α -chain domains raise important questions with respect to formation and regulation of convertases. In the series of activating and deactivating steps (from C3, C3b, iC3b to C3c), when do these rearrangements occur? How do the rearrangements relate to the binding and functioning of the various regulatory factors and receptors? Possibly, most rearrangements occur upon activation, in which case C3b and iC3b will resemble C3c and differ primarily in unravelling of the CUB domain. Alternatively, the rearrangements occur in a gradual process yielding different arrangements of each molecule. The C3 and C3c structures facilitate further structural and mutational studies of the activation and regulation of this central component of the complement system, and create new opportunities for drug development, targeting a wide variety of inflammatory diseases that have been associated with complement activation.

METHODS

Protein purification. C3c was purified using methods as described previously for C3 (ref. 42), with slight modifications. C3c from outdated human plasma (stored for several weeks at 4 °C) was purified by polyethylene glycol (PEG) precipitation, anion-exchange chromatography (DEAE Sephacel), cation-exchange chromatography (CM-Sephadex C50) and size-exclusion chromatography (Sephacryl 300). Trace amounts of IgG and IgA were removed by immune absorption. C3c was concentrated to 20 mg ml⁻¹ and dialysed against 10 mM Tris pH 7.4, 2 mM EDTA and 2 mM benzamidine. The glycan moiety on Asn 917 was cleaved off with N-glycosidase F (PNGase F) before crystallization. Both termini in the α -chain, generated by C3dg and C3f removal during plasma storage, displayed heterogeneity as observed by reduced gel and matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) analysis.

C3 was purified from frozen human plasma as described previously⁴², with slight modifications. C3 was purified by PEG precipitation, plasminogen depletion (Sephacrose 4B-L-lysine), anion-exchange chromatography (DEAE Sephacel) and size-exclusion chromatography (Sephacrose CL-6B). Trace amounts of IgG, IgA, IgM, C5 and factor H were removed by immune absorption. C3 was concentrated to 8 mg ml⁻¹ and precipitated by dialysis against 5 mM MES, pH 6.0 and stored at –80 °C until use to retain haemolytic activity. Before crystallization, C3 was solubilized by dialysis against 10 mM Tris, pH 7.4.

Crystallization and data collection. C3c was crystallized in hanging drops from mother liquor containing 18% w/v PEG-3000 and 200 mM LiNO₃ at 20 °C. Crystals grew to 100 × 100 × 100 μ m within 2 weeks. For cryo-protection 20%

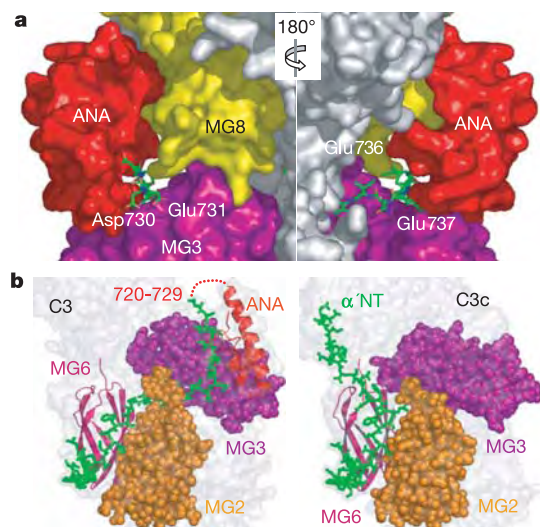


Figure 5 | The α' NT region (residues 727–744) slips through the β -ring. **a**, Two views, rotated by 180°, of the cone formed by ANA, MG3 and MG8 (surface representation), and residues 730–744 (stick representation). Residues important for factor B binding are labelled. **b**, The α' NT regions are on opposite sides of the molecule in C3 and C3c. In C3 this region is covalently linked to ANA; the scissile bond 726–727 is part of a disordered loop 720–729 (dashed line). Also shown are surface contours of C3 and C3c (transparent light grey), residues 727–767 (green sticks), ANA and MG6 (ribbons), and MG2 and MG3 of the β -ring (spheres).

v/v glycerol was added to the mother liquor, and crystals were flash-cooled in liquid nitrogen. Crystals displayed space group $P2_12_12$ ($a = 126.9$, $b = 246.9$, $c = 87.4$ Å), contained two molecules per asymmetric unit and diffracted to 2.4 Å resolution at ESRF beamline ID14-EH4. Varying pseudo-B centring was observed between crystals with extinctions for $(h + l) = \text{odd}$ reflections ranging from 10% to 100% (at 100%, crystals exhibited space group $B22_12$). In addition, crystals showed significant variations in cell dimensions. Diffraction data were processed using MOSFLM/CCP4 (ref. 17) and Denzo/Scapecap⁴³. Heavy-atom derivatives were prepared by soaking crystals in mother liquor containing 1–5 mM of heavy atom compounds for 24 h at 20 °C.

C3 was crystallized in hanging drops in 6% w/v PEG-monomethyl ether 550 and 100 mM sodium acetate at 20 °C. Crystals grew to $300 \times 100 \times 20$ µm within 2 weeks. Similar to C3c, 20% v/v glycerol was added as a cryo-protectant; crystals were flash-cooled in liquid nitrogen. Crystals exhibited space group $I222$ ($a = 117.0$, $b = 156.3$, $c = 271.2$ Å) with one molecule per asymmetric unit and diffracted to 3.3 Å resolution at ESRF beamline ID14-EH4. Diffraction data were processed with XDS and XSCALE⁴⁴.

Structure determination. Twenty mercury sites were found and refined in SOLVE⁴⁵ for a HgCl₂ derivative with diffraction data from a MAD experiment at mercury-inflection and high-remote energy. Twenty-seven platinum sites were found and refined in SOLVE⁴⁵ for a K₂PtCl₄ derivative from data of a SIRAS experiment at the platinum-peak energy and a data set collected from a native crystal. Phase combination and extension by multi-crystal and non-crystallographic symmetry averaging were crucial to produce an interpretable electron-density map. In the end, we used nine partial masks created in RAVE⁴⁶ to envelope a single molecule for averaging over six copies of the protein molecule in DMMULTI¹⁷. The resulting electron density was of good quality. A partial model was built automatically by ARP/wARP⁴⁷. The model of C3c was completed using O⁴⁸ and refined using REFMAC¹⁷ to a final R -value of 22.3% (R_{free} of 27.9%).

C3 was solved by molecular replacement. Initial positioning of C3c, or any of its fragments, was unsuccessful. However, the $\alpha 6$ – $\alpha 6$ barrel structure of C3d¹⁸ (18% of C3) was positioned correctly by Phaser¹⁹. Subsequently domains of C3c (C345C, MG2 plus MG6, MG5, MG4, MG8, MG3, MG7 and MG1, respectively) were added step-by-step using both Phaser¹⁹ and molecular graphics in O⁴⁸ with rigid-body refinement for positioning in Phaser. This yielded an 85% complete model that was completed by model building (of CUB, ANA and loops) and refined in CNS⁴⁹ and REFMAC¹⁷. The final refined model of C3 had R and R_{free} values of 23.6% and 29.5% (see Table 1; see also Supplementary Fig. 1 and Supplementary Table 1).

Received 8 June; accepted 5 July 2005.

- Carroll, M. C. The complement system in regulation of adaptive immunity. *Nature Immunol.* **5**, 981–986 (2004).
- Walport, M. J. Complement. First of two parts. *N. Engl. J. Med.* **344**, 1058–1066 (2001); Complement. Second of two parts. *N. Engl. J. Med.* **344**, 1140–1144 (2001).
- Sunyer, J. O., Zarkadis, I. K. & Lambris, J. D. Complement diversity: a mechanism for generating immune diversity? *Immunol. Today* **19**, 519–523 (1998).
- Levashina, E. A. *et al.* Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* **104**, 709–718 (2001).
- Budd, A., Blandin, S., Levashina, E. A. & Gibson, T. J. Bacterial $\alpha 2$ -macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol.* **5**, R38 (2004).
- Bokisch, V. A., Muller-Eberhard, H. J. & Cochrane, C. G. Isolation of a fragment (C3a) of the third component of human complement containing anaphylatoxin and chemotactic activity and description of an anaphylatoxin inactivator of human serum. *J. Exp. Med.* **129**, 1109–1130 (1969).
- Tack, B. F., Harrison, R. A., Janatova, J., Thomas, M. L. & Prahl, J. W. Evidence for presence of an internal thioester bond in third component of human complement. *Proc. Natl Acad. Sci. USA* **77**, 5764–5768 (1980).
- Lambris, J. D. The multifunctional role of C3, the third component of complement. *Immunol. Today* **9**, 387–393 (1988).
- Law, S. K., Lichtenberg, N. A. & Levine, R. P. Evidence for an ester linkage between the labile binding site of C3b and receptive surfaces. *J. Immunol.* **123**, 1388–1394 (1979).
- Muller-Eberhard, H. J. & Gotze, O. C3 proactivator convertase and its mode of action. *J. Exp. Med.* **135**, 1003–1008 (1972).
- Fishelson, Z., Pangburn, M. K. & Muller-Eberhard, H. J. Characterization of the initial C3 convertase of the alternative pathway of human complement. *J. Immunol.* **132**, 1430–1434 (1984).
- Pangburn, M. K., Schreiber, R. D. & Muller-Eberhard, H. J. Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein B1H for cleavage of C3b and C4b in solution. *J. Exp. Med.* **146**, 257–270 (1977).
- Ross, G. D., Lambris, J. D., Cain, J. A. & Newman, S. L. Generation of three different fragments of bound C3 with purified factor I or serum. I. Requirements for factor H vs CR1 cofactor activity. *J. Immunol.* **129**, 2051–2060 (1982).
- Seya, T., Turner, J. R. & Atkinson, J. P. Purification and characterization of a membrane protein (gp45-70) that is a cofactor for cleavage of C3b and C4b. *J. Exp. Med.* **163**, 837–855 (1986).
- Harrison, R. A. & Lachmann, P. J. Novel cleavage products of the third component of human complement. *Mol. Immunol.* **17**, 219–228 (1980).
- Lachmann, P. J., Pangburn, M. K. & Oldroyd, R. G. Breakdown of C3 after complement activation. Identification of a new fragment C3g, using monoclonal antibodies. *J. Exp. Med.* **156**, 205–216 (1982).
- Collaborative Computational Project, Number 4, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E. & Rini, J. M. X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science* **280**, 1277–1281 (1998).
- Storoni, L. C., McCoy, A. J. & Read, R. J. Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D* **60**, 432–438 (2004).
- de Bruijn, M. H. & Fey, G. H. Human complement component C3: cDNA coding sequence and derived primary structure. *Proc. Natl Acad. Sci. USA* **82**, 708–712 (1985).
- Huber, R., Scholze, H., Paques, E. P. & Deisenhofer, J. Crystal structure analysis and molecular model of human C3a anaphylatoxin. *Hoppe-Seyler's Z. Physiol. Chem.* **361**, 1389–1399 (1980).
- Jenner, L., Husted, L., Thirup, S., Sottrup-Jensen, L. & Nyborg, J. Crystal structure of the receptor-binding domain of $\alpha 2$ -macroglobulin. *Structure* **6**, 595–604 (1998).
- Thomas, M. L., Janatova, J., Gray, W. R. & Tack, B. F. Third component of human complement: localization of the internal thioester bond. *Proc. Natl Acad. Sci. USA* **79**, 1054–1058 (1982).
- Bramham, J. *et al.* Functional insights from the structure of the multifunctional C345C domain of C5 of complement. *J. Biol. Chem.* **280**, 10636–10645 (2005).
- Taniguchi-Sidle, A. & Isenman, D. E. Interactions of human complement component C3 with factor B and with complement receptors type 1 (CR1, CD35) and type 3 (CR3, CD11b/CD18) involve an acidic sequence at the N-terminus of C3 α '-chain. *J. Immunol.* **153**, 5285–5302 (1994).
- Sottrup-Jensen, L., Sand, O., Kristensen, L. & Fey, G. H. The α -macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian α -macroglobulins. *J. Biol. Chem.* **264**, 15781–15789 (1989).
- Lagueux, M., Perrodou, E., Levashina, E. A., Capovilla, M. & Hoffmann, J. A. Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of *Drosophila*. *Proc. Natl Acad. Sci. USA* **97**, 11427–11432 (2000).
- Ishii, N., Wadsworth, W. G., Stern, B. D., Culotti, J. G. & Hedgecock, E. M. UNC-6, a laminin-related protein, guides cell and pioneer axon migrations in *C. elegans*. *Neuron* **9**, 873–881 (1992).
- Sahu, A. & Lambris, J. D. Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunol. Rev.* **180**, 35–48 (2001).
- Pangburn, M. K., Schreiber, R. D. & Muller-Eberhard, H. J. Formation of the initial C3 convertase of the alternative complement pathway. Acquisition of C3b-like activities by spontaneous hydrolysis of the putative thioester in native C3. *J. Exp. Med.* **154**, 856–867 (1981).
- Isenman, D. E., Kells, D. I., Cooper, N. R., Muller-Eberhard, H. J. & Pangburn, M. K. Nucleophilic modification of human complement protein C3: correlation of conformational changes with acquisition of C3b-like functional properties. *Biochemistry* **20**, 4458–4467 (1981).
- Pangburn, M. K. Spontaneous reformation of the intramolecular thioester in complement protein C3 and low temperature capture of a conformational intermediate capable of reformation. *J. Biol. Chem.* **267**, 8584–8590 (1992).
- Law, S. K. & Dodds, A. W. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci.* **6**, 263–274 (1997).
- Gadjeva, M. *et al.* The covalent binding reaction of complement component C3. *J. Immunol.* **161**, 985–990 (1998).
- Dodds, A. W., Ren, X. D., Willis, A. C. & Law, S. K. The reaction mechanism of the internal thioester in the human complement component C4. *Nature* **379**, 177–179 (1996).
- van den Elsen, J. M. *et al.* X-ray crystal structure of the C4d fragment of human complement component C4. *J. Mol. Biol.* **322**, 1103–1115 (2002).
- Qazi, U., Gettins, P. G. & Stoops, J. K. On the structural changes of native human $\alpha 2$ -macroglobulin upon proteinase entrapment. Three-dimensional structure of the half-transformed molecule. *J. Biol. Chem.* **273**, 8987–8993 (1998).
- Daoudaki, M. E., Becherer, J. D. & Lambris, J. D. A 34-amino acid peptide of the third component of complement mediates properdin binding. *J. Immunol.* **140**, 1577–1580 (1988).
- Winters, M. S., Spellman, D. S. & Lambris, J. D. Solvent accessibility of native and hydrolyzed human complement protein 3 analyzed by hydrogen/deuterium

- exchange and mass spectrometry. *J. Immunol.* **174**, 3469–3474 (2005).
40. Oran, A. E. & Isenman, D. E. Identification of residues within the 727–767 segment of human complement component C3 important for its interaction with factor H and with complement receptor 1 (CR1, CD35). *J. Biol. Chem.* **274**, 5120–5130 (1999).
41. Lambris, J. D. *et al.* Dissection of CR1, factor H, membrane cofactor protein, and factor B binding and functional sites in the third complement component. *J. Immunol.* **156**, 4821–4832 (1996).
42. Hammer, C. H., Wirtz, G. H., Renfer, L., Gresham, H. D. & Tack, B. F. Large scale isolation of functionally active components of the human complement system. *J. Biol. Chem.* **256**, 3995–4006 (1981).
43. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
44. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
45. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
46. Kleywegt, G. J. & Jones, T. A. Software for handling macromolecular envelopes. *Acta Crystallogr. D* **55**, 941–944 (1999).
47. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
48. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
49. Brunger, A. T. *et al.* Crystallography NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
50. Solomon, K. R., Sharma, P., Chan, M., Morrison, P. T. & Finberg, R. W. CD109 represents a novel branch of the α 2-macroglobulin/complement gene family. *Gene* **327**, 171–183 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge the help of beamline scientists of the EMBL/ESRF and in particular R. B. G. Ravelli for help in data collection, and A. Perrakis and D. Egan for help and use of their crystallization robots. We thank A. T. Brunger, M. Bowen, B. DeLaBarre, J. D. Lambris, C. W. Vogel, D. Fritzing, T. Springer and T. K. Sixma for critically reading the manuscript. This work was supported by a 'Pionier' programme grant (P.G.) of the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (NWO-CW), the Dutch Kidney Foundation (A.R.), the Swedish Research Council (B.N.) and faculty grants at the University of Kalmar (K.N.-E.). A.R. does not support the evolutionary implications.

Author Information Co-ordinates and structure factors have been deposited in the Protein Data Bank under accession numbers 2A73 (C3) and 2A74 (C3c). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.G. (p.gros@chem.uu.nl).

ARTICLES

Evolutionary information for specifying a protein fold

Michael Socolich^{1,2*}, Steve W. Lockless^{1,2*†}, William P. Russ^{1,2}, Heather Lee^{1,2}, Kevin H. Gardner^{2,3}
& Rama Ranganathan^{1,2}

Classical studies show that for many proteins, the information required for specifying the tertiary structure is contained in the amino acid sequence. Here, we attempt to define the sequence rules for specifying a protein fold by computationally creating artificial protein sequences using only statistical information encoded in a multiple sequence alignment and no tertiary structure information. Experimental testing of libraries of artificial WW domain sequences shows that a simple statistical energy function capturing coevolution between amino acid residues is necessary and sufficient to specify sequences that fold into native structures. The artificial proteins show thermodynamic stabilities similar to natural WW domains, and structure determination of one artificial protein shows excellent agreement with the WW fold at atomic resolution. The relative simplicity of the information used for creating sequences suggests a marked reduction to the potential complexity of the protein-folding problem.

A fundamental tenet of biochemistry is that the amino acid sequence of a protein specifies its atomic structure and biochemical function¹. But exactly what information in the sequence of a protein is necessary and sufficient for producing the fold and its biological activity? Despite considerable progress in understanding the mechanisms of protein folding^{2,3}, the answer to this fundamental question remains unknown. The main problem is the vast potential complexity of cooperative interactions between amino acids—processes by which the free energy contribution of one residue depends on those of other residues^{4,5}. These amino acid couplings could be pairwise and local in the three-dimensional structure, but could also involve more complex cooperativities in which collections of residues interact through three-way or higher-order couplings^{6–8}. Given that protein structures are typically compact and well packed^{9,10}, proteins could be dense and complex networks of inter-atomic interactions, requiring specification of a great number of mutual constraints between amino acid positions to define the fold.

An approach to defining the architecture of amino acid interactions in proteins is suggested by an evolution-based method known as the statistical coupling analysis (SCA). This method postulates that regardless of spatial location or underlying mechanism, the conserved functional coupling of sites in a protein should drive their mutual coevolution^{11,12}. Given a sufficiently large and diverse multiple sequence alignment (MSA) of a protein family, the mutual dependencies should be evident in the conserved statistical correlations between amino acid distributions at sites. Application of the SCA in several different protein families reveals two general conclusions: (1) the global pattern of coevolutionary interactions is sparse, so that a small set of positions mutually coevolves among a majority that are largely decoupled, and (2) the strongly coevolving residues are spatially organized into physically connected networks linking distant functional sites in the structure through packing interactions^{12–15}. Studies involving directed mutagenesis^{12–14}, structure determination¹⁶, NMR dynamics¹⁷, computational modelling^{18,19} and literature study¹⁵ implicate these networks of

coevolving residues in contributing to core aspects of protein function.

More surprising, however, is the finding of sparseness. The SCA implies an unexpected degree of simplicity in amino acid interactions, with far fewer important constraints between residue pairs than would be expected from inspection of the atomic structure. Indeed, as has been pointed out^{20,21}, the evolution-based mapping of amino acid interactions does not look like the contact graph of the protein structure; many direct packing interactions show coevolution scores close to zero, and some distant sites linked through networks of coevolving residues are predicted to be coupled. Thus, the SCA mapping provides a picture of proteins as sparsely coupled architectures with redundant strong constraints linking a few sites, and a great deal of near-independent variation at most sites.

To test the overall hypothesis, we reasoned that if (and only if) the information contained in the SCA is a good estimate of the total sequence information for specifying a protein, it should be possible to computationally build artificial members of the protein family using no information except the SCA-based parameters of sequence conservation and coupling. In principle, these artificial sequences should fold into a structure representative of the family, and should function in a manner indistinguishable from their natural counterparts. In this and the accompanying paper²², we test this hypothesis in a computationally and experimentally facile model system, the WW domain.

SCA-based protein design

We began by carrying out the SCA for an alignment of 120 members of the WW domain family. WW domains are small, independently folding protein interaction modules that adopt a curved three-stranded β -sheet configuration and bind to proline-containing target sequences^{23,24} (Fig. 1a). In the SCA, conservation at each position in the MSA is given in an energy-like statistical parameter that measures the deviation of the observed distribution of amino acids from their mean values found in all proteins (Fig. 1b). The evolutionary

¹Howard Hughes Medical Institute and Departments of ²Pharmacology and ³Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9050, USA.

[†]Present address: The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

*These authors contributed equally to this work.

correlations between sequence positions are extracted by carrying out a perturbation analysis on the MSA; the conservation of amino acids at one site is perturbed (usually by restricting the site to one amino acid), and the impact of this perturbation on the conservation of amino acids at each of the other sites is measured in a distinct energy-like statistical parameter. Supplementary Fig. 1 shows an example of this calculation for one site in the WW MSA. Figure 1c shows a matrix representation of statistical coupling values for five different site-specific perturbations in the WW MSA that demonstrates the core results of the SCA. Perturbation experiments at all these sites simply expose the same redundant pattern of coevolution between a small set of moderately conserved positions. Not all moderately conserved positions are coupled—several other sites showing similar conservation scores (for example, positions 16, 17 or 25; Fig. 1b) show little coevolution with the sites chosen for perturbation.

The simplicity of these findings suggests the possibility that just the frequency distribution of amino acids at sites (conservation) plus the few rules of coupling in the SCA matrix may amount to the total constraints on WW sequences. To test this idea, we developed two computational algorithms for designing novel protein sequences using only SCA information. The first algorithm tests the necessity of the information in the SCA matrix by building artificial sequences that preserve the amino acid composition at sites but eliminate all statistical couplings between sites. Thus, this algorithm presumes that conservation can be adequately described as an intrinsic property of each site. To implement this algorithm, we simply select amino acids at each site independently based on the observed frequency distribution in the natural alignment. Accordingly, the SCA matrix for an alignment of 120 such sequences (termed the site-independent conservation (IC) model) shows no statistical coupling between positions (Fig. 1d). The second algorithm tests the sufficiency of the SCA matrix by building artificial sequences that preserve both the conservation pattern and the pattern of statistical couplings. These sequences are built using a Monte-Carlo-based simulated annealing protocol in which amino acids are completely shuffled within every column of the MSA while minimizing the difference between all coupling values in the SCA matrix for the artificial alignment and for the natural alignment. At convergence, this algorithm produces alignments of novel sequences ($\sim 10^5$ substitutions per sequence) that display the same conservation pattern and also closely reproduce the pattern of statistical couplings

observed in the natural alignment (termed coupled conservation (CC) model; compare Fig. 1c and e).

Construction of designed sequences

To evaluate the designed sequences for folding, we constructed libraries of synthetic genes for expression of the artificial proteins in *Escherichia coli*. Four libraries were built using a DNA-oligonucleotide-based gene synthesis protocol (Supplementary Fig. 2): (1) 42 natural WW sequences (N), drawn randomly from the MSA; (2) 43 IC sequences, built on the premise that conservation is strictly an intrinsic property of each site; (3) 43 CC sequences, built on the premise that conservation is a distributed property, parsed among sites in the manner described by the SCA matrix; and (4) 19 random sequences (R), in which amino acids at all sites were randomly drawn from their mean frequencies in the WW MSA. The natural sequences were built as positive controls because we do not know how many of these will fold when expressed as recombinant proteins in bacteria. The random sequences are negative controls; they contain no site-specific information and are not expected to fold.

Table 1 shows the comparative statistical properties of the designed sequences. Natural, IC and CC WW sequences show a mean amino acid identity to all natural WW domains of $\sim 36\%$, an expected result because all of these sequences contain the same pattern of conservation at sites. Another expected finding was that random sequences show much weaker identities to natural WW domains ($\sim 6\%$). However, the IC and CC sequences also show similar 'top-hit' identities—the per cent identity to their closest counterpart in the natural world ($55.7 \pm 5.6\%$ (IC) and $58.6 \pm 7.2\%$ (CC), $P = 0.11$, Kolmogorov–Smirnov test; mean $\pm$ s.d. is shown). This finding shows that despite additional constraints in their design, CC sequences are about as diverged as IC sequences; by this measure, the number of extra constraints arising from the SCA matrix is small.

Experimental analysis of designed sequences

Figure 2a shows a flowchart of experiments for each WW protein. Proteins were expressed as His₈-tagged fusions, purified using Ni²⁺-NTA affinity chromatography, and subjected to SDS–polyacrylamide gel electrophoresis (PAGE) analysis to evaluate expression and solubility. All libraries contained sequences that expressed poorly despite multiple attempts, or produced insoluble aggregates (Fig. 2b, first and second columns); these were scored as not folded. In

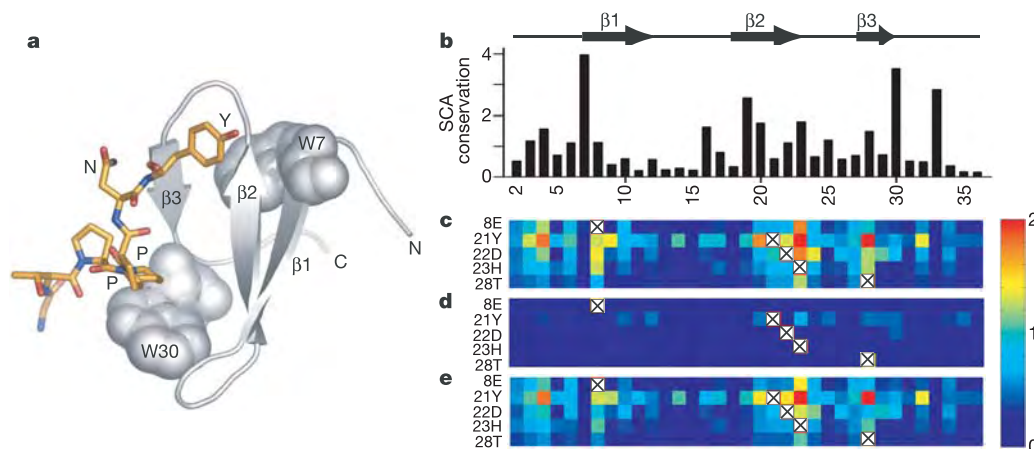


Figure 1 | SCA-based protein design. **a**, Structure of a representative WW domain (Nedd4.3, Protein Data Bank 1I5H) in complex with a target peptide (in stick representation). The two canonical tryptophans are shown as space-filling side chains. The figure was prepared using PyMol⁵¹. **b**, SCA conservation scores for each position in the WW alignment in arbitrary units of statistical energy¹². Position numbers (x axis) and the secondary structure diagram at the top coincide with matrix columns in **c–e**. **c**, A matrix representation of statistical coupling values from perturbation

analysis of five positions (rows) in the WW domain MSA. **d**, The matrix for an alignment of IC sequences, built by randomly selecting amino acids at each site from the observed frequency distributions in the natural alignment. **e**, The matrix for an alignment of CC sequences, derived from a design algorithm where both the conservation pattern and the pattern of statistical couplings in the natural alignment are preserved. Scale bar shows the SCA coevolution score, ranging from 0 (blue) to 2 (red).

Table 1 | Statistical properties of natural and artificial WW domains

WW library	Identity to natural WW domains in MSA (%; mean ± s.d.)	Identity to closest natural WW domain in MSA (%; mean ± s.d.)	Number of sequences
Natural* (folded)	35.3 ± 6.1 (36.2 ± 5.8)	80.4 ± 14.5 (78.2 ± 15.9)	42 (28)
IC†	36.1 ± 3.0	55.7 ± 5.6	43
CC‡ (folded)	35.0 ± 4.8 (37.4 ± 5.4)	58.6 ± 7.2 (63.1 ± 6.0)	43 (12)
Random§	6.3 ± 2.5	14.9 ± 4.3	19

* Drawn randomly from the natural WW MSA.
† Created by randomly selecting amino acids at each site from the observed frequency distribution at that site in the natural WW MSA.
‡ Created by Monte-Carlo simulation.
§ Created by randomly selecting amino acids at each site from their overall mean frequencies in the WW MSA.

addition, all libraries contained some fraction of well-expressed and soluble proteins (Fig. 2b, third column). These were further evaluated using thermal denaturation experiments and ¹H-NMR for evidence of a native state, as described below. Expression data for all 147 synthetic proteins are provided in Supplementary Fig. 3.

A hallmark of natively folded small proteins is cooperative and reversible transition between folded and unfolded states. In the WW domain, this folding equilibrium and the consistency of the two-state approximation have been well described^{25,26}. Here, we followed the folding reaction by monitoring the fluorescence of a buried tryptophan (Trp 7), which becomes quenched due to solvent exposure upon thermal denaturation and therefore reports the fraction of protein folded as a function of temperature (for example, Fig. 3a–c). We tested all 105 well-expressed and soluble proteins from the four sequence libraries for cooperative and reversible thermal transitions (Supplementary Fig. 4); a representative sampling of these data is shown in Fig. 3. Natural WW domains show a range of thermal denaturation profiles (Fig. 3a). Some, such as N1, are clearly well folded, showing a cooperative denaturation with thermodynamic parameters typical for WW domains (van't Hoff enthalpy (ΔH_u^{VH}) = 22.5 kcal mol^{−1}, T_m = 46.8 °C, where T_m is the melting temperature). Others, such as N22, cooperatively denature but are less stable (ΔH_u^{VH} = 18.5 kcal mol^{−1}, T_m = 25.2 °C). Still others, such as N36, despite good solubility, show no convincing evidence of a native state given the experimental conditions of the assay. To provide independent support for categorization of these denaturation profiles as folded or not folded, we collected ¹H-NMR spectra for a representative sampling of natural WW proteins. Characteristic features of folded WW proteins include good chemical shift dispersion of peaks corresponding to backbone amide protons

(δ > 9 parts per million (p.p.m.)), often accompanied by distinct chemical shifts of the two indole nitrogen protons down-field of 10 p.p.m. (the two canonical Trp residues are in distinct chemical environments—one in the core and one solvent exposed, see Fig. 1a), and up-field chemical shifts (δ < 0.5 p.p.m.) corresponding to side-chain methyl protons²⁶. Spectra for N1 (Fig. 3d) and N22 (Fig. 3e) confirm that these are indeed natively folded. As suggested by thermal melts, the NMR spectrum for N36 (Fig. 3f) shows none of these hallmarks, consistent with an unfolded protein. On the basis of the analysis of natural WW sequences, we determined rigorous criteria for folding: artificial sequences were declared as folded only if they showed cooperative and reversible thermal denaturation and ¹H-NMR spectra consistent with a native state (Supplementary Figs 4 and 5).

Figure 3b, g–i shows similar data for a representative set of CC sequences. Some, such as CC16 (Fig. 3i), are, like N36, soluble but not folded by either criterion. Others, such as CC45 (Fig. 3g) or CC18 (Fig. 3h), show thermal denaturation profiles that fall into the same range as natural WW domains (CC45: ΔH_u^{VH} = 32.4 kcal mol^{−1}, T_m = 65.6 °C; CC18: ΔH_u^{VH} = 19.63 kcal mol^{−1}, T_m = 34.3 °C), and show strong evidence in ¹H-NMR spectra of being natively folded. In contrast, no IC sequences showed evidence of native folding. Figure 3c shows thermal denaturation profiles for every soluble IC sequence (n = 30, grey) and Fig. 3j–l shows corresponding ¹H-NMR spectra for the three IC sequences that constitute the best case scenarios by thermal melts. The data demonstrate that some CC sequences are natively folded with thermodynamic properties similar to that of natural WW domains, and that no IC sequences are natively folded.

Figure 4a summarizes the experimental analysis of all 147 WW

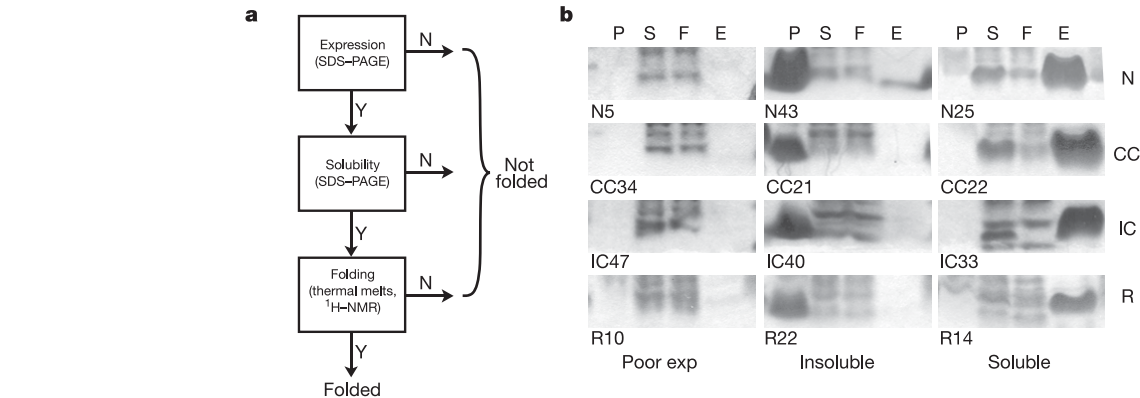


Figure 2 | Experimental evaluation of artificial proteins (part 1). **a**, A flowchart of experiments for evaluating WW sequences. Each sequence was expressed as a His₈-tagged fusion in *E. coli*, tested for solubility, and if soluble, subjected to thermal denaturation and, in most cases, ¹H-NMR. **b**, SDS-PAGE analysis of WW sequences drawn from four different sequence libraries: N, natural; CC, coupled conservation model; IC, site-independent conservation model; R, random. Lanes represent pellet or insoluble fraction (P), soluble fraction (S), flow-through after Ni²⁺-NTA affinity chromatography (F) and eluted proteins (E).

sequences tested. Sixty-seven per cent of natural WW sequences were folded by the criteria described above, a number that simply reflects the efficiency of producing randomly chosen WW domains in our expression system. As expected, no random sequences were folded, although nearly half of these were well expressed and soluble. No IC sequences were natively folded, although a substantial fraction (70%) was soluble. The finding that these sequences showed a greater probability of being soluble in comparison to random sequences suggests that conservation taken at sites alone may provide enough information about the folding process to permit hydrophobic collapse to a molten globule-like state. However, the site-independent conservation model is insufficient to produce the native state. In contrast, over one-quarter of the CC sequences were natively folded. Figure 4b compares the unfolding enthalpy and melting temperatures derived from thermal denaturation experiments for all folded natural WW domains and CC sequences, showing that the CC sequences fall into the same range of thermodynamic parameters as their natural counterparts ($P = 0.1$, multivariate analysis of variance (ANOVA)). Just like natural WW domains, the CC domains are marginally stable folds in which a fine balance of opposing forces probably accounts for the distinction between folded and non-folded states.

Folding to a native state involves efficient packing of hydrophobic atoms within the interior of proteins, a factor that leads to higher average sequence conservation in the core of proteins²⁷. Although CC sequences show similar divergence from natural WW domains as IC sequences overall (Table 1), we wondered whether CC sequences

might natively fold because they are more similar to WW domains within the core. To examine this, we calculated sequence identities within core residues (3, 7, 20, 22, 33) between CC or IC sequences and natural WW domains. CC sequences show $66.7 \pm 7.0\%$ mean and $97 \pm 7.1\%$ top-hit identities to natural sequences for these positions (mean $\pm$ s.d.), values that reflect the near invariance of some core positions in the MSA. However, IC sequences are no different from CC sequences, showing $68.3 \pm 9\%$ mean and $95.4 \pm 9.4\%$ top-hit identities to natural sequences for core positions (mean $\pm$ s.d.). Also, the folded subset of CC sequences shows the same core sequence identity ($67.8 \pm 7\%$ mean and $98.3 \pm 6\%$ top-hit; mean $\pm$ s.d.). Thus, native folding in CC sequences is not explained by a more natural-like composition of core residues.

Taken together, the data argue that in addition to the amino acid distributions at sites, the statistical coupling information is necessary and sufficient to specify native folding. Because no information about the WW domain except the coupling values in the SCA matrix was used to constrain the CC sequences beyond the IC sequences, we conclude that the specific topology of mutual constraints between sites predicted by the SCA is one solution for achieving the folded state of this protein family.

Atomic structure of an artificial WW domain

Do the artificial proteins adopt the canonical WW fold? To examine this, we pursued in-depth structural analysis of one of the CC proteins, CC45. CC45 shows 39% mean identity and 61% top-hit identity to natural WW sequences, values that are near the average

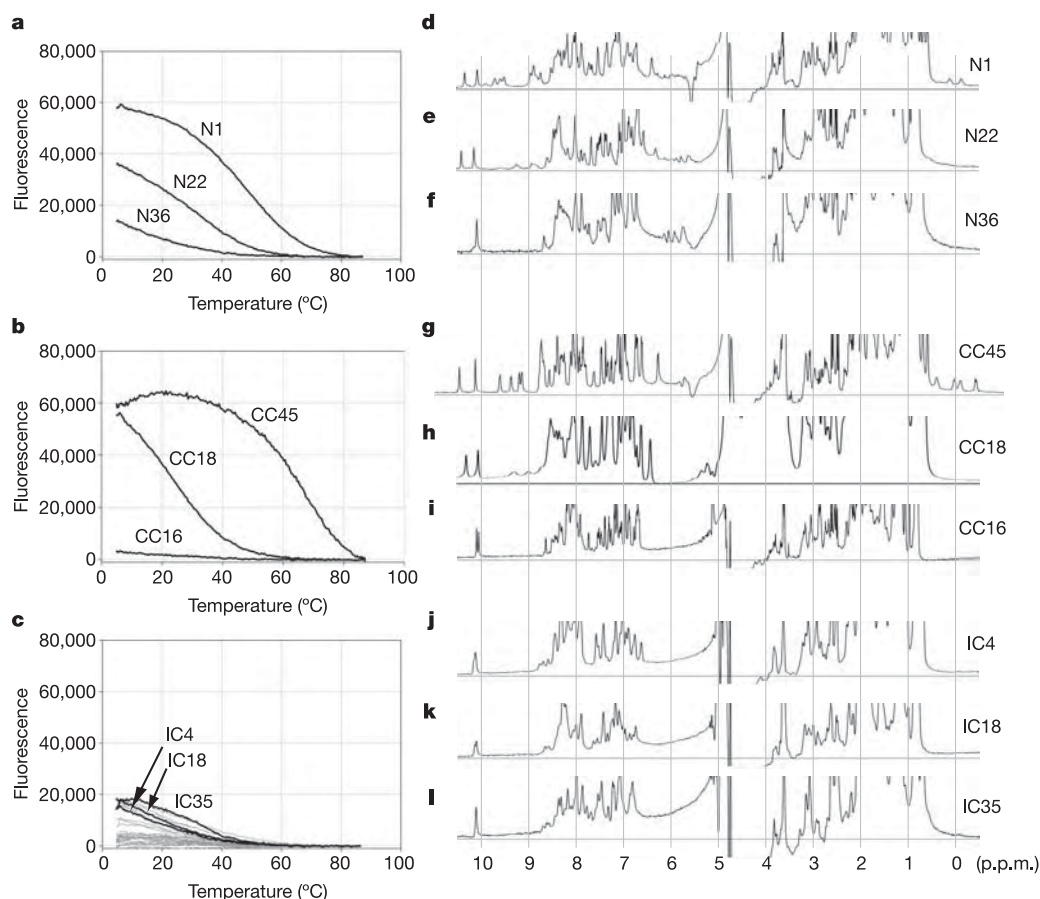


Figure 3 | Experimental evaluation of artificial proteins (part 2). Thermal denaturation studies and corresponding ^1H -NMR spectra for a sampling of WW sequences drawn from the natural (a, d–f), CC (b, g–i), or IC model (c, j–l) alignments. For natural and CC sequences, three sequences are shown that were highly stable (d, g), moderately stable (e, h), and unfolded

(f, i). For IC sequences, melts of every soluble sequence ($N = 30$) are overlaid in grey; ^1H -NMR spectra of three sequences (in black) with the best potential for some native structure are shown in j–l. Whereas natural and CC sequences include some that are folded, no IC sequences are folded. Fluorescence is reported in counts s^{-1} .

for both CC and IC sequences (Table 1). As described, this protein is well folded by several independent biophysical assays (Fig. 3b, g). We solved the three-dimensional structure of CC45 by solution NMR methods, using 800 distance and dihedral angle restraints (Fig. 5a, b; see also Supplementary Table 2). CC45 was refined to reasonably high precision, clearly confirming the curved three-stranded antiparallel β -sheet structure characteristic of all WW domains. However, the structural similarity of CC45 to other WW domains seems to go well beyond just the fold level. Tertiary structure motifs common to WW domains are found in CC45, including a centrally located tryptophan (W7) that sits upon a platform of two proline side chains (P4, P33, Fig. 5b). In addition, several sites in CC45 display unusual proton chemical shifts based on comparison with the BioMagRes-Bank database of protein NMR data (for example, $\delta = -0.38$ p.p.m. for N22 H β 2, Fig. 3g; see Supplementary Table 1 for further examples). Such unusual shifts arise from unique tertiary packing that places protons in close proximity to aromatic side chains. Comparison of the chemical shifts of CC45 to those of two natural WW domains, Pin1 (ref. 28) and Nedd4.3 (ref. 29), showed that all three proteins display these same unusual shifts at analogous positions in each sequence. Thus, CC45 adopts a stable WW-like three dimensional structure.

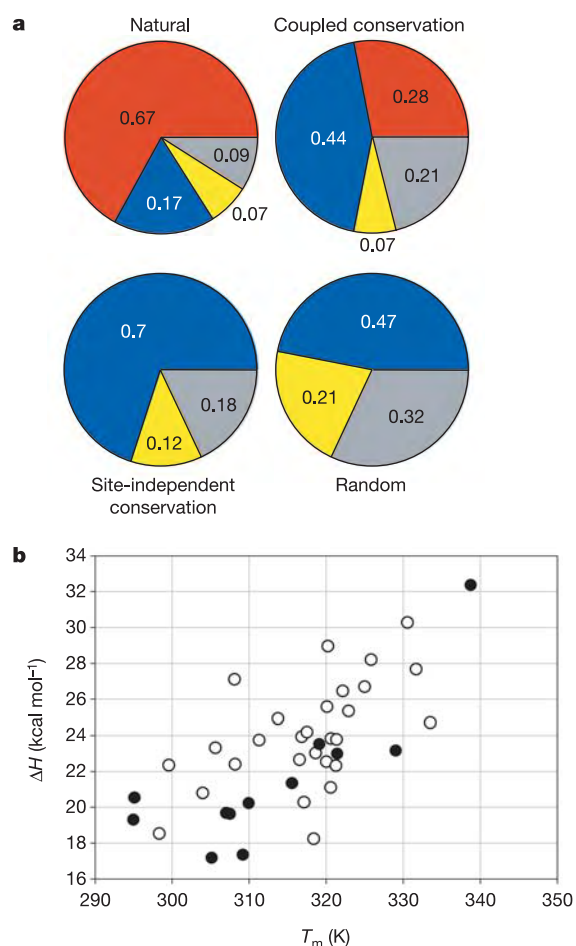


Figure 4 | Summary of experiments on all natural and artificial WW sequences. **a**, A pie chart showing the outcomes of folding studies for natural ($n = 42$), CC ($n = 43$), IC ($n = 43$), or random ($n = 19$) WW sequences. Red, natively folded; blue, soluble but unfolded; yellow, insoluble; grey, poor expressing. **b**, Melting temperatures (T_m) and van't Hoff enthalpies of unfolding for all folded WW sequences. Open circles indicate natural sequences and filled circles indicate the 12 folded CC sequences. The artificial sequences show thermodynamic parameters that fall into the same range as that of natural WW domains.

To examine how well the SCA-based design recapitulates the native structure of the WW domain, we overlaid the structure of CC45 with those of several different natural WW domains by minimizing the root mean squared deviations (r.m.s.d.) of backbone C α atoms for all structures (Fig. 5c). All seven structures clearly adopt very similar backbone folds, and no obvious feature seems to distinguish CC45 from the other proteins. To quantify this result, we compared the average pairwise r.m.s.d. values for backbone atoms of all the natural WW domains (1.52 ± 0.4 Å) with the r.m.s.d. values of CC45 from all the natural domains (1.19 ± 0.65 Å). The difference between these distributions is not significant ($P = 0.34$), demonstrating that CC45 is as similar at atomic resolution to natural WW domains as natural domains are to each other.

Conclusions

Classical studies indicate that protein structure and function results from globally minimizing the free energy of the polypeptide chain under physiological conditions¹. In this work, we have tested a model for the specific pattern of amino acid interactions that makes up the global free energy minimum using the WW domain as a model system (Fig. 1b, c). The model is based on three core hypotheses: (1) that amino acid interactions specifying the atomic structure are conserved throughout members of a protein family rather than being idiosyncratic; (2) that conservation is a distributed rather than site-independent property because it fundamentally arises from the cooperativity of energetic interactions; and (3) that the parsing of conservation is well estimated by the statistical energy function contained in the SCA method. The finding that a significant fraction of CC sequences are natively folded whereas IC sequences are not provides strong support for these hypotheses. This result is particularly informative because CC sequences are statistically indistinguishable from IC sequences with regard to sequence divergence from natural WW domains. We conclude that it is the specific distribution of conservation rather than the quantity of conservation that dictates native folding.

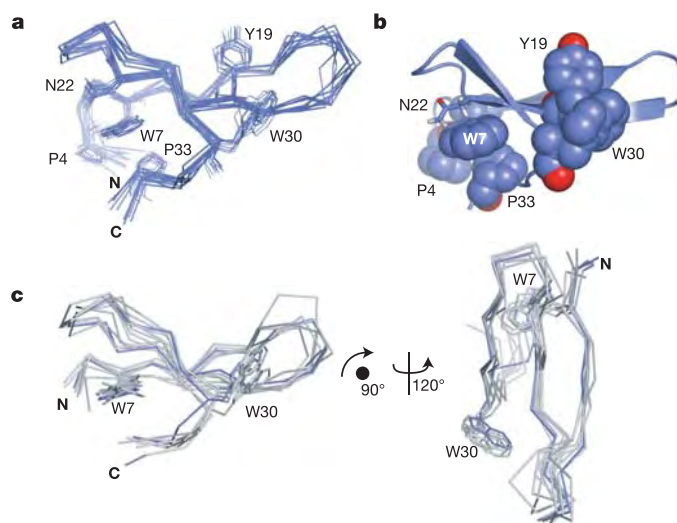


Figure 5 | NMR structure determination of CC45, an artificial WW domain. **a**, Ensemble of ten lowest energy structures determined for CC45, showing backbone traces of each structure and side chains of selected residues discussed. **b**, Ribbon diagram of the representative structure of the CC45 ensemble. Residues highlighted in CPK or stick bonds illustrate packing interactions in the core (Trp 7, Pro 33, Asn 22), and in the canonical proline-binding pocket (Tyr 19, Trp 30). **c**, Structure-based alignment of CC45 (blue) with six natural WW domains in white (FBP28WW (1E0L), YJQ8WW (1E0N), dystrophin (1EG3), Nedd4.3 (1I5H), YAP65 (1K9R) and Pin1 (1PIN)). The data show that the CC45 adopts a WW-like tertiary structure.

How is the sparse architecture of residue interactions suggested by the SCA consistent with the fact that proteins are well packed, and with the intuitive sense that residues closer to each other in the structure should interact more strongly? Indeed, mutations made in spatially local sites ($<5 \text{ \AA}$ apart) are more likely to be thermodynamically coupled³⁰, a finding that has been used to argue that the energetic architecture of proteins could be dense and local rather than sparse and distributed²¹. However, mutation-based free energy measurements do not measure the energetic value of wild-type residues; they measure the energetic effect due to mutation. When close in space, it becomes difficult to separate the average spatial correlation of mutation-induced effects in proteins from the intrinsic interactions between the wild-type residues in proteins. For example, alanine mutations are often thought of as loss-of-function mutations in proteins, but these mutations can cause local cavities to form in structures, and the energetic value of such mutations includes a cavity-dependent contribution³¹. In such cases, it is not surprising that pairs of mutations in local regions show energetic coupling. In contrast, SCA-based protein design is a massive perturbation experiment in which the mutagenesis at every site is constrained by its own frequency distribution in the MSA and by coupling to those few other sites that evolution shows can provide compensatory mutations. Consequently, a large number of packing interactions are decoupled in the design of CC sequences and vary independently, consistent with findings that proteins are generally tolerant to mutation³². We suggest that this experiment provides a more informative and extensive test of the architecture of amino acid interactions in proteins.

Nevertheless, it is also clear that the density of atoms within the core of proteins is high, similar to that in crystals of free amino acids^{9,10}. At first glance, this finding seems to highlight the importance of local interactions in proteins to create this high level of order. However, recent studies show that although the mean density of atoms in the core of proteins suggests a solid-like arrangement, the distribution of packing densities is much more like that of a liquid^{33,34}. That is, packing is heterogeneous in protein cores with some highly ordered regions mixed with many regions that are not as highly ordered. A number of studies are now beginning to add experimental support for the heterogeneous and distributed nature of mechanical interactions in proteins^{35–37}. It will be important to test further the physical model that emerges from these studies and the SCA: proteins are heterogeneous energetic architectures with a network of coupled residues existing in the environment of many weaker interactions.

METHODS

Statistical coupling analysis. Statistical coupling analysis was performed as previously described¹². WW domain sequences were collected using PSI-BLAST³⁸ (e -scores ≤ 0.001) from the non-redundant database of protein sequences (release date October 2000), and aligned using ClustalW³⁹. The five statistical perturbations used in this study (8E, 21Y, 22D, 23H and 28T) were chosen by knowledge of structural or functional importance in the WW domain^{40,41}.

Protein design algorithms. IC model sequences were created by randomly drawing amino acids for each site from the corresponding amino acid frequency distributions in the natural WW alignment. Random sequences were created by randomly selecting residues at every site from the overall frequency distribution in the MSA. CC model sequences were generated using a Monte-Carlo simulated annealing algorithm, described in the Supplementary Methods section.

Gene construction and protein expression. Genes encoding artificial sequences were designed by back-translating designed protein sequences using *E. coli* codon optimization (Vector NTI Suite, Informax Inc), built on overlapping DNA oligonucleotides, assembled using the polymerase chain reaction, and cloned into the pHIS8-3 expression vector (provided by J. Noel). Constructs were verified by DNA sequencing. Proteins were expressed in JM109(DE3) cells grown at 37 °C in Terrific broth to an absorbance at 600 nm of ~ 1.2 , and induced with 0.5 mM IPTG at 18 °C overnight. For fluorescence experiments, 50 ml culture was lysed in 3 ml binding buffer (25 mM TrisHCl, pH 8.0, 0.5 M NaCl, 5 mM imidazole) by sonication followed by centrifugation and incubation of the

cleared lysate with 75 μl bed volume Ni^{2+} -NTA (Qiagen) for 30 min at 4 °C. After washing (three times with 15 ml binding buffer), WW proteins were eluted in elution buffer (100 mM TrisHCl, pH 8.0, 1 M NaCl, 400 mM imidazole). Cultures were scaled up for NMR experiments, using 1–2 l of Terrific broth and 0.5 ml of affinity resin.

Thermal denaturation assay. Purified proteins were diluted into 2 ml elution buffer for a final WW domain concentration of ~ 1 – $10 \mu\text{M}$. Protein fluorescence was monitored at 340 nm (excitation at 295 nm) using a spectrofluorometer (Photon Technologies Inc.) outfitted with a Peltier temperature controller. Data were acquired while changing the temperature from 4 °C to 90 °C and back to 4 °C, at a rate of 2°C min^{-1} with 5 min equilibration periods at 4 °C and 90 °C. Melting curves for 10 μM free tryptophan were subtracted to account for the intrinsic temperature dependence of Trp fluorescence. Melting temperature (T_m) and enthalpy of unfolding at the T_m were calculated by fitting the first derivative of the denaturation curves to the differential form of the van't Hoff equation⁴².

NMR spectroscopy. All NMR spectra were recorded on 500 and 600 MHz Varian Inova spectrometers. One-dimensional ^1H -NMR spectra of various WW domains were obtained on samples containing 100 μM to 1 mM protein in 100 mM NaCl, 100 mM sodium phosphate buffer (pH 7.0) and 90%:10% H_2O : D_2O . All spectra were recorded at 25 °C using a 3–9–19 Watergate sequence⁴³ for water suppression. For the CC45 chemical shift assignments and solution structure determination, a combination of two-dimensional double-quantum filtered correlation spectroscopy (DQF-COSY), total correlated spectroscopy (TOCSY; $\tau_{\text{mix}} = 60 \text{ ms}$) and nuclear Overhauser effect spectroscopy (NOESY; $\tau_{\text{mix}} = 150 \text{ ms}$) were recorded on 600 μM unlabelled protein samples in the above buffer at 25 °C and 38 °C, along with two-dimensional TOCSY ($\tau_{\text{mix}} = 60 \text{ ms}$) and NOESY ($\tau_{\text{mix}} = 60, 150 \text{ ms}$) spectra recorded at 25 °C on a 1 mM protein sample made in the same buffer with 99% D_2O solvent. Additional spectra were recorded on an 850 μM sample of uniformly ^{15}N -labelled CC45 in 90%:10% H_2O : D_2O , including three-dimensional ^{15}N -edited TOCSY ($\tau_{\text{mix}} = 60 \text{ ms}$), three-dimensional ^{15}N -edited NOESY ($\tau_{\text{mix}} = 150 \text{ ms}$), three-dimensional HNHA and two-dimensional $^{15}\text{N}/^1\text{H}$ heteronuclear single quantum correlation (HSQC) experiments. All spectra were processed using the NMRPipe package⁴⁴ and analysed using nmrView⁴⁵. The CC45 sequence is $\text{NH}_3^+ \text{-MPLPPGWERTDVEGKVYYFNVRLTLTTWERPTIILE-COO}^-$.

Structure determination. Distance restraints were obtained by analysing the three-dimensional ^{15}N -edited NOESY and two-dimensional homonuclear NOESY spectra recorded in 99% D_2O (both $\tau_{\text{mix}} = 150 \text{ ms}$) using the CNS package⁴⁶ with the ARIA 1.2 extension⁴⁷. Additional restraints on the backbone dihedral angles ϕ and ψ were derived from a TALOS-based analysis of backbone chemical shifts⁴⁸ and supplemented by 3J(HN-H α) coupling constants measured in a three-dimensional HNHA spectrum. Hydrogen bond restraints for a total of six backbone amides were generated for sites showing β -sheet structure by dihedral angle analysis and typical inter-strand nuclear Overhauser effects (NOEs). One hundred structures were generated in the last ARIA iteration, out of which the ten with the lowest energies were selected for a final refinement stage in water. The resulting ensemble of ten structures was used for all analyses as implemented in the programs PROCHECK-NMR⁴⁹ and MOLMOL⁵⁰. Ramachandran statistics for CC45 are: $80.7 \pm 6.3\%$ in most favoured regions, $18.0 \pm 7.3\%$ in additionally allowed regions, $1.0 \pm 1.6\%$ in generously allowed regions, and $0.3 \pm 1.0\%$ in disfavoured regions.

Received 25 January; accepted 30 June 2005.

1. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
2. Daggett, V. & Fersht, A. The present view of the mechanism of protein folding. *Nature Rev. Mol. Cell Biol.* **4**, 497–502 (2003).
3. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. & Karplus, M. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**, 331–339 (2000).
4. Hidalgo, P. & MacKinnon, R. Revealing the architecture of a K^+ channel pore through mutant cycles with a peptide inhibitor. *Science* **268**, 307–310 (1995).
5. Horovitz, A. & Fersht, A. R. Co-operative interactions during protein folding. *J. Mol. Biol.* **224**, 733–740 (1992).
6. Chen, J. & Stites, W. E. Higher-order packing interactions in triple and quadruple mutants of staphylococcal nuclease. *Biochemistry* **40**, 14012–14019 (2001).
7. LiCata, V. J. & Ackers, G. K. Long-range, small magnitude nonadditivity of mutational effects in proteins. *Biochemistry* **34**, 3133–3139 (1995).
8. Luque, I., Leavitt, S. A. & Freire, E. The linkage between protein folding and functional cooperativity: two sides of the same coin? *Annu. Rev. Biophys. Biomol. Struct.* **31**, 235–256 (2002).

9. Gerstein, M. & Chothia, C. Packing at the protein-water interface. *Proc. Natl Acad. Sci. USA* **93**, 10167–10172 (1996).
10. Richards, F. M. & Lim, W. A. An analysis of packing in the protein folding problem. *Q. Rev. Biophys.* **26**, 423–498 (1993).
11. Lichtarge, O., Bourne, H. R. & Cohen, F. E. An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* **257**, 342–358 (1996).
12. Lockless, S. W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299 (1999).
13. Hatley, M. E., Lockless, S. W., Gibson, S. K., Gilman, A. G. & Ranganathan, R. Allosteric determinants in guanine nucleotide-binding proteins. *Proc. Natl Acad. Sci. USA* **100**, 14445–14450 (2003).
14. Shulman, A. I., Larson, C., Mangelsdorf, D. J. & Ranganathan, R. Structural determinants of allosteric ligand activation in RXR heterodimers. *Cell* **116**, 417–429 (2004).
15. Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Struct. Biol.* **10**, 59–69 (2003).
16. Peterson, F. C., Penkert, R. R., Volkman, B. F. & Prehoda, K. E. Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition. *Mol. Cell* **13**, 665–676 (2004).
17. Fuentes, E. J., Der, C. J. & Lee, A. L. Ligand-dependent dynamics and intramolecular signalling in a PDZ domain. *J. Mol. Biol.* **335**, 1105–1115 (2004).
18. Estabrook, R. A. *et al.* Statistical coevolution analysis and molecular dynamics: identification of amino acid pairs essential for catalysis. *Proc. Natl Acad. Sci. USA* **102**, 994–999 (2005).
19. Ota, N. & Agard, D. A. Intramolecular signalling pathways revealed by modeling anisotropic thermal diffusion. *J. Mol. Biol.* **351**, 345–354 (2005).
20. Fodor, A. A. & Aldrich, R. W. Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins* **56**, 211–221 (2004).
21. Fodor, A. A. & Aldrich, R. W. On evolutionary conservation of thermodynamic coupling in proteins. *J. Biol. Chem.* **279**, 19046–19050 (2004).
22. Russ, W. P., Lowery, D. M., Mishra, P., Yaffe, M. B. & Ranganathan, R. Natural-like function in artificial WW domains. *Nature* doi:10.1038/nature03990 (this issue).
23. Macias, M. J. *et al.* Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide. *Nature* **382**, 646–649 (1996).
24. Bork, P. & Sudol, M. The WW domain: a signalling site in dystrophin? *Trends Biochem. Sci.* **19**, 531–533 (1994).
25. Jager, M., Nguyen, H., Crane, J. C., Kelly, J. W. & Gruebele, M. The folding mechanism of a β -sheet: the WW domain. *J. Mol. Biol.* **311**, 373–393 (2001).
26. Koepf, E. K., Petrassi, H. M., Sudol, M. & Kelly, J. W. WW: An isolated three-stranded antiparallel β -sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Sci.* **8**, 841–853 (1999).
27. Bashford, D., Chothia, C. & Lesk, A. M. Determinants of a protein fold. Unique features of the globin amino acid sequences. *J. Mol. Biol.* **196**, 199–216 (1987).
28. Wintjens, R. *et al.* 1H NMR study on the binding of Pin1 Trp-Trp domain with phosphothreonine peptides. *J. Biol. Chem.* **276**, 25150–25156 (2001).
29. Kanelis, V., Rotin, D. & Forman-Kay, J. D. Solution structure of a Nedd4 WW domain-ENaC peptide complex. *Nature Struct. Biol.* **8**, 407–412 (2001).
30. Schreiber, G. & Fersht, A. R. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478–486 (1995).
31. Eriksson, A. E. *et al.* Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178–183 (1992).
32. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* **247**, 1306–1310 (1990).
33. Liang, J. & Dill, K. A. Are proteins well-packed? *Biophys. J.* **81**, 751–766 (2001).
34. Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128–132 (2005).
35. Benkovic, S. J. & Hammes-Schiffer, S. A perspective on enzyme catalysis. *Science* **301**, 1196–1202 (2003).
36. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295**, 1520–1523 (2002).
37. Frauenfelder, H., McMahon, B. H. & Fenimore, P. W. Myoglobin: the hydrogen atom of biology and a paradigm of complexity. *Proc. Natl Acad. Sci. USA* **100**, 8615–8617 (2003).
38. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
39. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
40. Huang, X. *et al.* Structure of a WW domain containing fragment of dystrophin in complex with β -dystroglycan. *Nature Struct. Biol.* **7**, 634–638 (2000).
41. Kananov, J., Pirozzi, G., Uveges, A. J. & Kay, B. K. Characterizing Class I WW domains defines key specificity determinants and generates mutant domains with novel specificities. *Chem. Biol.* **8**, 231–241 (2001).
42. John, D. M. & Weeks, K. M. van't Hoff enthalpies without baselines. *Protein Sci.* **9**, 1416–1419 (2000).
43. Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. Gradient-tailored water suppression for ^1H - ^{15}N HSQC experiments optimized to retain full sensitivity. *J. Magn. Reson. A* **102**, 241–245 (1993).
44. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
45. Johnson, B. A. & Blevins, R. A. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
46. Brünger, A. T. *et al.* Crystallography and NMR system (CNS): a new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
47. Linge, J. P., O'Donoghue, S. I. & Nilges, M. *Methods in Enzymology* 71–90 (Academic Press, 2001).
48. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
49. Laskowski, R. A., Rullman, J. A. C., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).
50. Koradi, R., Billeter, M. & Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).
51. Delano, W. L. *The PyMOL Molecular Graphics System* (<http://www.pymol.org>) (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. G. Gilman, M. Rosen, and members of the Ranganathan laboratory for advice and critical review of the manuscript, and E. Olson for contributing to this project. R.R. and S.W.L. thank the students of the 2001 Woods Hole physiology course for participating in an initial phase of this work. This study was supported by the Robert A. Welch foundation (R.R. and K.H.G.), the Mallinckrodt Foundation Scholar Award (R.R.), and the NIH (K.H.G.). W.P.R. is an associate and R.R. is an investigator of the Howard Hughes Medical Institute.

Author Information Atomic coordinates for CC45 have been deposited in the Protein Data Bank under accession code 1YMZ. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.R. (rama.ranganathan@utsouthwestern.edu).

A large population of galaxies 9 to 12 billion years back in the history of the Universe

O. Le Fèvre¹, S. Paltani¹, S. Arnouts¹, S. Charlot^{2,3}, S. Foucaud⁴, O. Ilbert^{1,5}, H. J. McCracken^{3,12}, G. Zamorani⁶, D. Bottini⁴, B. Garilli⁴, V. Le Brun¹, D. Maccagni⁴, J. P. Picat⁷, R. Scaramella⁸, M. Scodeggio⁴, L. Tresse¹, G. Vettolani⁸, A. Zanichelli⁸, C. Adami¹, S. Bardelli⁶, M. Bolzonella⁵, A. Cappi⁶, P. Ciliegi⁶, T. Contini⁷, P. Franzetti⁴, I. Gavaud^{7,9}, L. Guzzo¹⁰, A. Iovino¹⁰, B. Marano⁵, C. Marinoni¹⁰, A. Mazure¹, B. Meneux¹, R. Merighi⁶, R. Pellò⁷, A. Pollo¹⁰, L. Pozzetti⁶, M. Radovich¹¹, E. Zucca⁶, M. Arnaboldi¹¹, M. Bondi⁸, A. Bongiorno⁵, G. Busarello¹¹, L. Gregorini⁸, F. Lamareille⁷, G. Mathez⁷, Y. Mellier^{3,12}, P. Merluzzi¹¹, V. Ripepi¹¹ & D. Rizzo⁷

To understand the evolution of galaxies, we need to know as accurately as possible how many galaxies were present in the Universe at different epochs¹. Galaxies in the young Universe have hitherto mainly been identified using their expected optical colours^{2–4}, but this leaves open the possibility that a significant population remains undetected because their colours are the result of a complex mix of stars, gas, dust or active galactic nuclei. Here we report the results of a flux-limited I-band survey of galaxies at look-back times of 9 to 12 billion years. We find 970 galaxies with spectroscopic redshifts between 1.4 and 5. This population is 1.6 to 6.2 times larger than previous estimates^{2–4}, with the difference increasing towards brighter magnitudes. Strong ultraviolet continua (in the rest frame of the galaxies) indicate vigorous star formation rates of more than 10–100 solar masses per year. As a consequence, the cosmic star formation rate representing the volume-averaged production of stars is higher than previously measured at redshifts of 3 to 4.

One technique to identify galaxies in the distant Universe is to use the discontinuity in the spectrum produced by the photon absorption below the 912 Å Lyman limit to isolate high-redshift galaxies from the dominating population of foreground galaxies in colour-colour diagrams. This technique has led to the identification of the Lyman-break galaxies (LBGs) at redshifts $z \approx 2–4$ (refs 2–4), the current reference for the high-redshift population. However, this colour-selection technique with its associated uncertainties and assumptions could exclude some significant fraction of the high-redshift galaxy population. The question therefore naturally arises whether there are galaxies at similar distances that have escaped searches at optical wavelengths. This can be checked using the basic observational approach of selecting all galaxies in a region of the sky brighter than a given apparent flux limit, followed by a systematic spectroscopic redshift measurement. This magnitude-selection technique has been used successfully from $z \approx 0$ up to $z \approx 1.5$ (refs 5–9), but assembling significant samples of a few hundred $z > 2$ galaxies requires spectroscopic observations of several thousands of sources.

Here we report the spectroscopic identification of galaxies with $1.4 \leq z \leq 5$ from a purely I-band ($\sim 8,100$ Å) flux-selected sample. The VIMOS VLT (Very Large Telescope) Deep Survey (VVDS) has assembled a ‘first-epoch’ sample of $\sim 8,000$ galaxies in the VVDS field in the Cetus constellation with measured redshifts in the range

$0 \leq z \leq 5$ (ref. 10). There are 970 galaxies with $1.4 \leq z \leq 5$; the spectra clearly show all the features that are expected in this rest-frame wavelength range (Fig. 1). Using the ultraviolet continuum flux at 1,500 Å rest wavelength (L_{1500} , in $\text{erg s}^{-1} \text{Hz}^{-1}$), and the conversion¹¹ to star formation rate (SFR, in solar masses per year, $M_{\odot} \text{yr}^{-1}$) given by $\text{SFR} = 1.4 \times 10^{-28} L_{1500}$, we find that these galaxies have $\text{SFR} = (10–100) M_{\odot} \text{yr}^{-1}$, without correction for dust absorption, indicating a strong star formation activity. Interestingly, only a small fraction ($\sim 15\%$) of these high-redshift spectra show Ly α in emission, with an average Ly α equivalent width of only 3 Å, a property already observed in other samples¹².

We estimate from these data the number density of galaxies in the redshift range $1.4 \leq z \leq 5$ as the number of galaxies corrected for the fraction of galaxies we have observed in the field, $\sim 27\%$, and for the estimated fraction of galaxies with incorrect redshifts (Fig. 1), divided by the observed area, which is 1,700 arcmin². The galaxy surface density is reported in Table 1 for each redshift range, and presented in Fig. 2. An additional uncertainty in these estimates is due to the large-scale fluctuations in the distribution of galaxies. We checked the amplitude of these variations in the VVDS field by computing surface densities in sub-fields of sizes from 850 down to 100 arcmin². We find differences of up to a factor of 3 in the smaller fields, but only about 3% in the larger fields; these are smaller than the quoted statistical errors of our density measurements and consistent with simulation predictions¹³. Importantly, the number density values reported here can be regarded as conservative lower limits, as the galaxies that have the right colours to be LBG candidates³, but for which we could not assign a redshift, were discarded. If all of these objects are at the redshift suggested by their colours, they could increase the projected galaxy density by additional factors of 1.5 to 2 at the fainter magnitudes at $z \approx 3$ and $z \approx 4$, respectively. In addition, we do not correct our density estimates for those galaxies which, while being in the redshift ranges considered here, may have been wrongly classified at lower redshift. With this conservative approach, we find a total surface density of galaxies down to a magnitude $I_{AB} = 24$ of $\Sigma(z = [1.4, 2]) = 0.758 \pm 0.101$, $\Sigma(z = [2.7, 3.4]) = 0.235 \pm 0.025$ and $\Sigma(z = [3.7, 4.5]) = 0.063 \pm 0.020$ galaxies arcmin⁻², using the same redshift ranges as in previous studies for comparison^{3,4}. The galaxy surface densities we measure in the VVDS at $z \approx 3$ are between

¹Laboratoire d'Astrophysique de Marseille, OAMP, UMR6110, CNRS, Université de Provence Aix-Marseille I, BP8, F-13376 Marseille cedex 12, France. ²Max Planck Institut für Astrophysik, D-85741 Garching, Germany. ³Institut d'Astrophysique de Paris, UMR 7095, 98 bis Bvd Arago, F-75014 Paris, France. ⁴IASF-INAF, via Bassini 15, I-20133, Milano, Italy. ⁵Università di Bologna, Dipartimento di Astronomia, via Ranzani 1, I-40127, Bologna, Italy. ⁶INAF, Osservatorio Astronomico di Bologna, via Ranzani 1, I-40127, Bologna, Italy. ⁷Laboratoire d'Astrophysique de l'Observatoire Midi-Pyrénées (UMR 5572), 14 avenue E. Belin, F-31400 Toulouse, France. ⁸IRA-INAF, via Gobetti 101, I-40129, Bologna, Italy. ⁹European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany. ¹⁰INAF, Osservatorio Astronomico di Brera, via Brera 28, Milan, Italy. ¹¹INAF, Osservatorio Astronomico di Capodimonte, via Moiariello 16, I-80131, Napoli, Italy. ¹²Observatoire de Paris, LERMA, 61 Avenue de l'Observatoire, F-75014 Paris, France.

1.6 and 6.2 times larger than those obtained from the colour-selected Lyman-break samples³.

This result therefore shows that the Universe contained more galaxies at redshifts 1.4 to 5 than previously reported using the colour-selection technique. The probability that the VVDS galaxy surface densities are consistent with the previously reported surface density of LBGs is less than 10^{-4} for the whole sample. Remarkably, we find that in the R-band magnitude intervals $22.5 \leq R \leq 23.0$, $23.0 \leq R \leq 23.5$ and $23.5 \leq R \leq 24.0$, the VVDS galaxy surface density is higher than that of colour-selected LBGs by factors of 6.2, 2.1 and 1.6 respectively, with corresponding probabilities of only 0.3%, 0.2% and 4.3% that the surface densities from the two samples are consistent with each other. This indicates that the bright end of the galaxy luminosity function at these early epochs is more populated than previously thought. The VVDS galaxies in the (u - g - r) diagram are mainly located near the boundary used to isolate LBGs in the colour-selection technique, as shown in Fig. 3. A significant fraction of galaxies has relatively red g - r colours, which may indicate the contribution of a reddened active galactic nucleus (AGN) to this population. Steidel *et al.*³ claim that the fraction of $z \approx 3$ galaxies that erroneously escape the LBG selection at magnitude $I_{AB} = 24$ is about 40%, a figure highly dependent on the assumptions about the properties of the high-redshift population. Our results demonstrate that magnitude-selected samples, although requiring a large number of redshift measurements, provide a more complete census of the high-redshift galaxy population. More details about the properties of

Table 1 | Projected galaxy density in the VVDS

I_{AB} magnitude	$z = [1.4, 2]$	$z = [2.7, 3.4]$	$z = [3.7, 4.5]$
22.0–22.5	0.021 ± 0.006	0.005 ± 0.002	0.001 ± 0.001
22.5–23.0	0.084 ± 0.013	0.027 ± 0.005	0.005 ± 0.003
23.0–23.5	0.227 ± 0.033	0.060 ± 0.008	0.013 ± 0.005
23.5–24.0	0.433 ± 0.061	0.147 ± 0.016	0.049 ± 0.016

The number of galaxies per square arc minute in 0.5-magnitude intervals is tabulated for 3 redshift intervals from $z = 1.4$ to $z = 4.5$. The galaxy densities have been corrected for the sampling rate of the VVDS and the estimated fraction of incorrect redshifts.

this new population of bright high-redshift galaxies (BHZGs) will be given elsewhere (S.P. *et al.*, manuscript in preparation).

While the VVDS provides direct evidence for higher galaxy densities at large redshifts than previously thought, this may have already been indirectly hinted at. From photometric-redshift analysis, Pascarelle *et al.*¹⁴ reported finding more galaxies outside the LBG colour box than expected, although their result is weakened by their large photometric error bars; Foucaud *et al.*¹⁵ reported a large surface density of LBGs from colour-colour selection, which is difficult to interpret as due to contamination by galaxies at other redshifts. An interesting result is the report of very bright LBGs at $z \approx 3$ discovered in the Sloan Digital Sky Survey¹⁶; although this sample is small and might be partially contaminated by AGN, our extrapolated counts to bright magnitudes are consistent with the number density of the SDSS objects of 1.2×10^{-6} galaxies arcmin⁻² at magnitude $I_{AB} = 20$. Other searches of LBGs or Ly α emitters have produced samples of faint galaxies in fields about 20 times smaller, with little overlap with our luminosity range¹⁷, making a comparison with our sample difficult. The galaxies we have identified may only be partly related to dim red galaxies¹⁸ as they have a surface density 6 to 20 times smaller than the VVDS galaxies, and are dimmer and redder than galaxies in our survey. We also estimate from redshift surveys of the faint sub-millimetre population that sub-millimetre galaxies are about 10 times less numerous¹⁹, and are therefore mainly a different

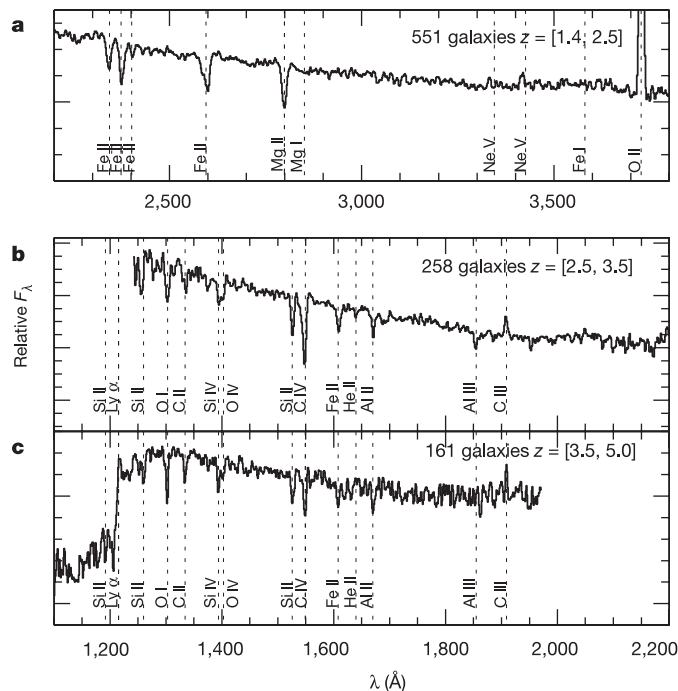


Figure 1 | Average spectra of VVDS galaxies. The average spectra of VVDS galaxies in three redshift ranges (**a**, $1.4 \leq z < 2.5$; **b**, $2.5 \leq z < 3.5$; and **c**, $3.5 \leq z < 5.0$). The main spectral features expected at these wavelengths are clearly visible in all spectra. There are 532 galaxies with reliable redshift measurements (quality flags 2, 3 and 4; ref. 10), and 438 of lower quality (flag 1; ref. 10). Since there are a number of uncertain redshift determinations in the VVDS, we must estimate the contamination of our sample by galaxies outside the redshift range of interest, which have been erroneously included in this range during our redshift measurement process. From the comparison of the sum of stellar absorption equivalent widths in the stacked spectra of galaxies with low-quality redshift measurements to those in the stacked spectra of galaxies with secure redshifts¹⁰, we estimate that approximately 69%, 52% and 46% of all VVDS redshifts in the range $z = [1.4, 2]$, $[2.7, 3.4]$ and $[3.7, 4.5]$, respectively, must be correct. F_λ , flux at wavelength λ .

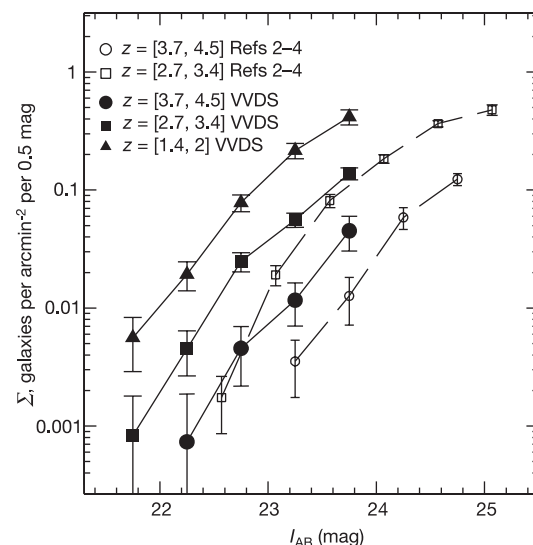


Figure 2 | Galaxy surface density. The VVDS galaxy surface density is computed in 3 redshift domains, $z = [1.4, 2]$ (filled triangles), $z = [2.7, 3.4]$ (filled squares) and $z = [3.7, 4.5]$ (filled circles), compared to the surface density measured by Steidel *et al.*^{2–4} (open symbols and dashed lines). The VVDS surface density is consistently larger by factors 1.6–6.2 at $z \approx 3$, and by factors 2–3.5 at $z \approx 4$. The surface densities plotted in this figure have been corrected for the survey sampling rate and the estimated fraction of incorrect redshifts. Errors include both the Poisson error associated with the observed number of galaxies and the errors in the estimated fraction of incorrect redshift measurements.

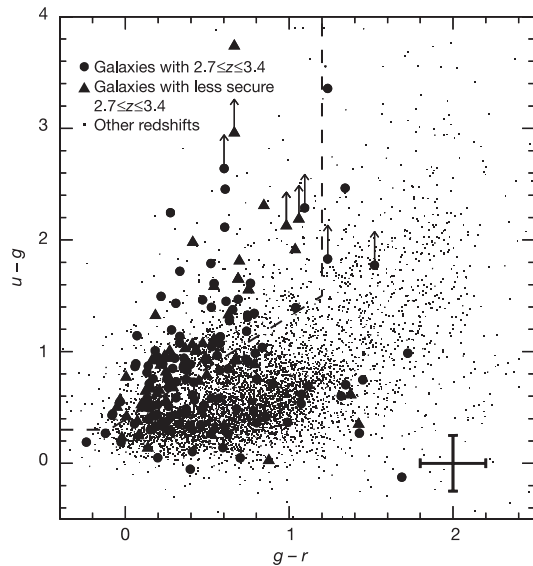


Figure 3 | The $(u-g, g-r)$ colour diagram of the VVDS high-redshift galaxies. The $u-g$ and $g-r$ colours of galaxies are computed from their magnitudes in the u , g and r bands (ultraviolet, green and red light, respectively). Galaxies with $2.7 \leq z \leq 3.4$ are represented by circle symbols, and galaxies with less reliable redshifts in the same redshift range are represented by triangles. Galaxies at other redshifts are represented by small dots. The dashed line delineates the area where LBGs with $2.7 \leq z \leq 3.4$ would be searched for if a colour-colour pre-selection was performed. The typical standard deviation errors on colours for the faintest galaxies are shown at the bottom right.

population. We conclude that we have observed a population of bright galaxies that has not been identified before.

Our discovery of a significantly larger galaxy population at high redshift than was previously believed has important consequences for our understanding of galaxy evolution. The galaxy formation process has to spawn a more numerous galaxy population producing more stars than previously assumed at a time when the Universe was ~ 10 – 20% of its current age. Summing the ultraviolet luminosity at $1,700 \text{ \AA}$ of galaxies brighter than a rest-frame absolute magnitude $M_{\text{AB}}(1,700 \text{ \AA}) = -22$ in the observed volume (without correction for interstellar dust absorption), we find an ultraviolet luminosity density $\text{LD}(1700 \text{ \AA}) = 1.6 \times 10^{25} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ Mpc}^{-3}$ and $1.4 \times 10^{25} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ Mpc}^{-3}$ at $z \approx 3$ and $z \approx 4$, respectively (for the concordance cosmology $\Omega = 0.3$, $\Lambda = 0.7$ and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$). This is more than twice that quoted in previous studies^{3,20,21}. The total number of stars formed in bright galaxies at this epoch may therefore be two to three times higher than previously inferred.

Received 26 April; accepted 29 June 2005.

1. White, S. D. M. & Frenk, C. S. Galaxy formation through hierarchical clustering. *Astrophys. J.* **379**, 52–79 (1991).

2. Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M. & Adelberger, K. L. Spectroscopic confirmation of a population of normal star-forming galaxies at redshifts $z > 3$. *Astrophys. J.* **462**, L17–L21 (1996).
3. Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M. & Pettini, M. Lyman-break galaxies at $z > 4$ and the evolution of the ultraviolet luminosity density at high redshift. *Astrophys. J.* **519**, 1–17 (1999).
4. Steidel, C. C. et al. A survey of star-forming galaxies in the $1.4 < z < 2.5$ redshift desert: overview. *Astrophys. J.* **604**, 534–550 (2004).
5. Glazebrook, K. et al. A faint galaxy redshift survey to $B = 24$. *Mon. Not. R. Astron. Soc.* **273**, 157–168 (1995).
6. Lilly, S. J., Le Fèvre, O., Crampton, D., Hammer, F. & Tresse, L. The Canada-France Redshift Survey. I. Introduction to the survey, photometric catalogs, and surface brightness selection effects. *Astrophys. J.* **455**, 50–59 (1995).
7. Le Fèvre, O., Crampton, D., Lilly, S. J., Hammer, F. & Tresse, L. The Canada-France Redshift Survey. II. Spectroscopic program: Data for the 0000–00 and 1000 + 25 Fields. *Astrophys. J.* **455**, 60–74 (1995).
8. Cowie, L. L., Hu, E. M. & Songaila, A. Detection of massive forming galaxies at redshifts $z > 1$. *Nature* **377**, 603–605 (1995).
9. Cimatti, A. et al. The K20 survey. I. Disentangling old and dusty star-forming galaxies in the ERO population. *Astron. Astrophys.* **381**, 68–72 (2002).
10. Le Fèvre, O. et al. The VIMOS VLT Deep Survey – First epoch VVDS-deep survey: 11564 spectra with $17.5 \leq I_{\text{AB}} \leq 24$, and the redshift distribution over $0 < z \leq 5$. *Astron. Astrophys.* (in the press); preprint at (<http://arxiv.org/astro-ph/0409133>) (2004).
11. Kennicutt, R. S. Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36**, 189–232 (1998).
12. Ando, M. et al. Lyman break galaxies at $z \sim 5$: Rest-frame ultraviolet spectra. *Astrophys. J.* **610**, 635–641 (2004).
13. Somerville, R. et al. Cosmic variance in the Great Observatories Origins Deep Survey. *Astrophys. J.* **600**, L71–L74 (2004).
14. Pascarella, S. M., Lanzetta, K. M. & Fernández-Soto, A. The ultraviolet luminosity density of the Universe from photometric redshifts of galaxies in the Hubble deep field. *Astrophys. J.* **508**, L1–L4 (1998).
15. Foucaud, S. et al. The Canada-France deep fields survey-II: Lyman-break galaxies and galaxy clustering at $z \sim 3$. *Astron. Astrophys.* **409**, 835–850 (2003).
16. Bentz, M. C., Osmer, P. S. & Weinberg, D. H. Bright Lyman break galaxy candidates in the Sloan Digital Sky Survey first data release. *Astrophys. J.* **600**, L19–L22 (2004).
17. Lehnert, M. & Bremer, M. Luminous Lyman break galaxies at $z > 5$ and the source of reionization. *Astrophys. J.* **593**, 630–639 (2003).
18. Förster Schreiber, N. M. et al. A substantial population of red galaxies at $z > 2$: Modeling of the spectral energy distributions of an extended sample. *Astrophys. J.* **616**, 40–62 (2004).
19. Chapman, S. C., Blain, A. W., Ivison, R. J. & Smail, I. A median redshift of 2.4 for galaxies bright at submillimetre wavelengths. *Nature* **422**, 695–698 (2003).
20. Giavalisco, M. et al. The rest-frame ultraviolet luminosity density of star-forming galaxies at redshifts $z > 3.5$. *Astrophys. J.* **600**, 103–106 (2004).
21. Bunker, A. et al. The star formation rate of the Universe at $z \sim 6$ from the Hubble Ultra-Deep Field. *Mon. Not. R. Astron. Soc.* **355**, 374–384 (2004).

Acknowledgements We thank the ESO for continuous support of this programme; the CNRS, the University of Provence and the Italian INAF for funding; and S. J. Lilly and A. Renzini for discussions. The observations reported here are based on observations obtained at the European Southern Observatory Very Large Telescope, and on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to O.L.F. (Olivier.LeFevre@oamp.fr).

LETTERS

An organic thyristor

F. Sawano¹, I. Terasaki¹, H. Mori^{2,3}, T. Mori⁴, M. Watanabe⁵, N. Ikeda⁶, Y. Nogami^{3,7} & Y. Noda⁵

Thyristors are a class of nonlinear electronic device that exhibit bistable resistance—that is, they can be switched between two different conductance states¹. Thyristors are widely used as inverters (direct to alternating current converters) and for the smooth control of power in a variety of applications such as motors and refrigerators. Materials and structures that exhibit nonlinear resistance of this sort are not only useful for practical applications: they also provide systems for exploring fundamental aspects of solid-state and statistical physics. Here we report the discovery of a giant nonlinear resistance effect in the conducting organic salt θ -(BEDT-TTF)₂CsCo(SCN)₄, the voltage-current characteristics of which are essentially the same as those of a conventional thyristor. This intrinsic organic thyristor works as an inverter, generating an alternating current when a static direct-current voltage is applied. Whereas conventional thyristors consist of a series of diodes (their nonlinearity comes from interface effects at the p-n junctions), the present salt exhibits giant nonlinear resistance as a bulk phenomenon. We attribute the origin of this effect to the current-induced melting of insulating charge-order domains, an intrinsically non-equilibrium phenomenon in the sense that ordered domains are melted by a steady flow.

The organic salt θ -(BEDT-TTF)₂CsCo(SCN)₄ is known as a charge-ordered conductor². BEDT-TTF stands for bis(ethylene-dithio)-tetrathiafulvalene, whose molecular structure is schematically shown in the inset of Fig. 1a. This salt is a layered material composed of conducting BEDT-TTF layers and insulating CsCo(SCN)₄ layers alternately stacked along the *b* axis. The Greek letter θ specifies the packing pattern of BEDT-TTF molecules in the conducting layer, representing a 'triangular lattice' in this case. On the triangular lattice of BEDT-TTF, one hole exists for every two molecules, which tends to be localized at low temperatures owing to Coulomb repulsion, hence the charge order^{3,4}. In this particular salt, different patterns of charge order compete^{5–7} and freeze inhomogeneously without any thermodynamic transitions^{8,9}. Previously¹⁰, we discovered giant nonlinear resistance in the related organic salt θ -(BEDT-TTF)₂CsZn(SCN)₄. The *b*-axis conductance abruptly jumped at a threshold electric field to a value 100 times larger, but its microscopic origin was left to be explored. Here we address its origin and report on the discovery of a unique phenomenon, the generation of an alternating current (a.c.) from a static direct-current (d.c.) voltage in a bulk single crystal.

Two single crystals (samples B1 and B2) were prepared using a galvanostatic anodic oxidation method. The detailed growth conditions and their characterization are described elsewhere². The dimensions of B1 and B2 were $0.19 \times 0.09 \times 0.35 \text{ mm}^3$ and $0.19 \times 0.07 \times 0.52 \text{ mm}^3$, respectively, with the shortest dimension along the *b* axis (the interlayer direction). Using these, we can easily convert the resistance, current and voltage into resistivity, current density and electric field. However, for the sake of clear understanding, we choose to show only raw values in the present paper,

because these are tightly bound to the measurement conditions, as will be shown later.

The voltage-current characteristics were measured by a two-probe method in series with a metal-oxide film resistor of known resistance (standard resistance R_{std} , see Fig. 1c) from 4.2 to 20 K in a liquid He cryostat. The contact resistance was 10–50 Ω at room temperature (300 K), more than 100 times smaller than the sample resistance below 20 K. The a.c. response was recorded by a digital oscilloscope equipped with a numerical fast Fourier transform (FFT) function.

Figure 1a shows the nonlinear resistance $V_{\text{sample}}/I_{\text{ex}}$, that is, the

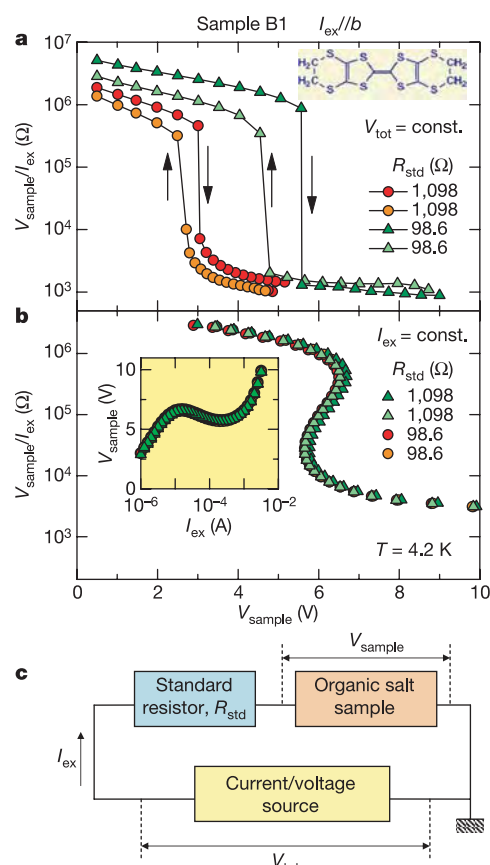


Figure 1 | Nonlinear resistance $V_{\text{sample}}/I_{\text{ex}}$ of a θ -(BEDT-TTF)₂CsCo(SCN)₄ crystal (sample B1) at 4.2 K. The external current I_{ex} is along the *b* axis (interlayer direction). **a**, $V_{\text{sample}}/I_{\text{ex}}$ measured in constant V_{tot} across the sample with a standard resistance R_{std} in series. A schematic drawing of the BEDT-TTF molecule is shown in the inset. **b**, $V_{\text{sample}}/I_{\text{ex}}$ measured in constant external current I_{ex} . **c**, Schematic figure of the measurement configuration.

¹Department of Applied Physics, Waseda University, Tokyo 169-8555, Japan. ²The Institute for Solid State Physics, The University of Tokyo, Kashiwa 277-8581, Japan. ³CREST, Japan Science and Technology Corporation, Kawaguchi 332-0012, Japan. ⁴Department of Organic and Polymeric Materials, Tokyo Institute of Technology, Tokyo 152-8552, Japan. ⁵Institute of Multidisciplinary Research for Advanced Materials, Tohoku University, Sendai 980-8577, Japan. ⁶Japan Synchrotron Radiation Research Institute, Spring-8, Mikazuki, Hyogo 679-5198, Japan. ⁷The Graduate School of Natural Science and Technology, Okayama University, Okayama 700-8530, Japan.

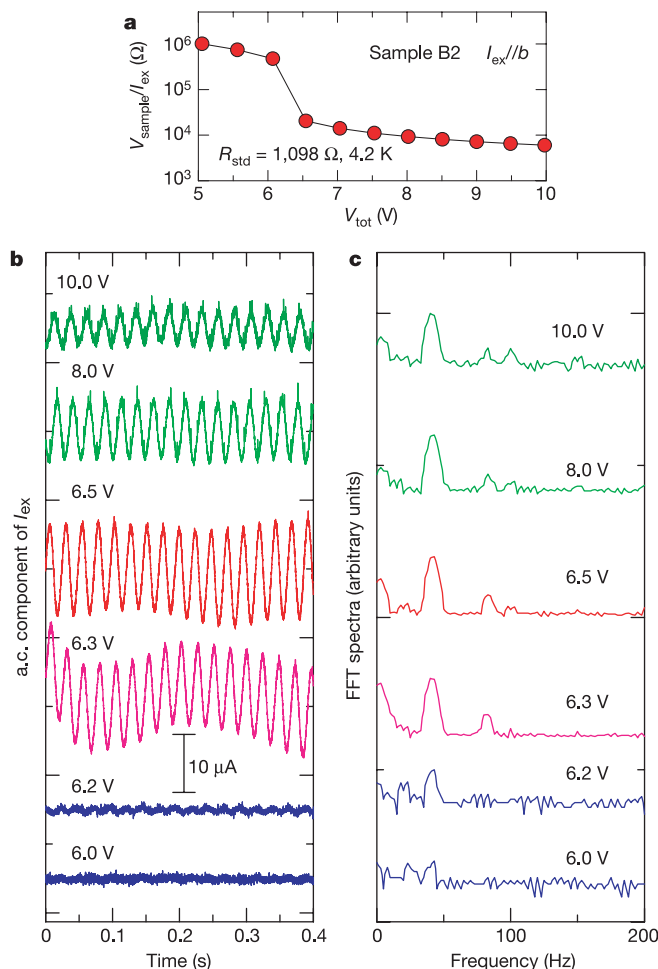


Figure 2 | Inverter (d.c.-a.c. conversion) phenomena in θ -(BEDT-TTF) $_2$ CsCo(SCN) $_4$ measured at 4.2 K. **a, Nonlinear resistance $V_{\text{sample}}/I_{\text{ex}}$ of the θ -(BEDT-TTF) $_2$ CsCo(SCN) $_4$ crystal (sample B1) along the b -axis direction as a function of V_{tot} . **b**, The a.c. voltage component across R_{std} for various V_{tot} . **c**, FFT spectrum of **b**.**

ratio of the voltage across the sample V_{sample} to the external current I_{ex} . The data were taken at 4.2 K in constant V_{tot} conditions (see Fig. 1c), and are plotted as a function of V_{sample} . Similarly to the CsZn salt¹⁰, $V_{\text{sample}}/I_{\text{ex}}$ shows a drastic jump at a certain voltage as well as hysteresis. Clearly, the threshold voltage depends on R_{std} , indicating that neither the threshold voltage nor the threshold electric field is intrinsic to the sample. In contrast, as shown in Fig. 1b, $V_{\text{sample}}/I_{\text{ex}}$ in constant I_{ex} conditions is perfectly independent of R_{std} , which strongly suggests that the constant I_{ex} conditions capture the intrinsic properties. As shown in the inset of Fig. 1b, sample B1 shows a sharp negative derivative resistance [$d(V_{\text{sample}}/I_{\text{ex}})/dI_{\text{ex}} < 0$] for $1 \times 10^{-5} \text{ A} < I_{\text{ex}} < 2 \times 10^{-4} \text{ A}$. The $V_{\text{sample}}-I_{\text{ex}}$ characteristics are essentially the same as those of thyristors, indicating that the present organic salt works as an organic thyristor—an intrinsic electronics component.

Surprisingly, sample B1 has also proved to work as a d.c.-a.c. inverter. Using the same set-up as in Fig. 1c, the a.c. voltage component across the standard resistor ($R_{\text{std}} = 1,098 \Omega$) was measured in various d.c. V_{tot} values at 4.2 K. As shown in Fig. 2a, the nonlinear resistance shows an abrupt jump between $V_{\text{tot}} = 6.2$ and 6.5 V. In exactly the same V_{tot} range, a large a.c. signal of 40 Hz suddenly appears (see Fig. 2b). The observed oscillation is nearly sinusoidal and has no significant higher harmonic components except for the second-order one, as shown in the FFT in Fig. 2c. Because Fig. 1c is an equivalent set-up for current-noise

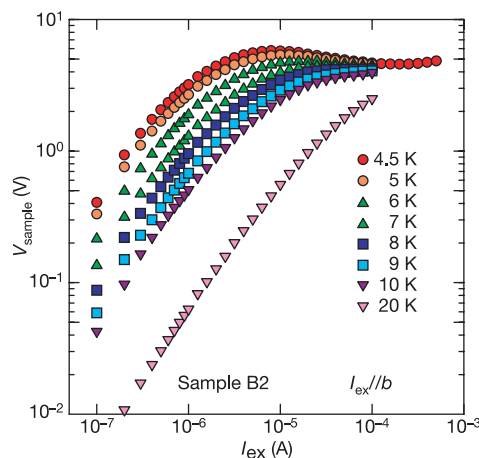


Figure 3 | Voltage-current characteristics of a second θ -(BEDT-TTF) $_2$ CsCo(SCN) $_4$ crystal (sample B2) in constant external current I_{ex} at various temperatures.

measurements, the data in Fig. 2b might be called noise, but what we observed here is not noise; in this case a 40-Hz oscillation is spontaneously induced by a constant d.c. flow, as found in manufactured inverter devices.

We can understand the inverter mechanism in terms of the observed nonlinear resistance as follows. Solutions of I_{ex} for a given V_{tot} are determined graphically by an intersection of $V_{\text{sample}} = V_{\text{tot}} - R_{\text{std}}I_{\text{ex}}$ and $V_{\text{sample}} = f(I_{\text{ex}})$ in the $I_{\text{ex}}-V_{\text{sample}}$ plane, where $f(I_{\text{ex}})$ is given by the inset of Fig. 1b. I_{ex} can take three solutions (one is unstable, the other two are stable) for a certain range of V_{tot} . In addition, we expect a capacitive component arising from the large dielectric constant¹⁰, which shifts the phase of the time-dependent current to enable the current density oscillate in between the two stable solutions¹¹. A similar oscillation has been observed in the Gunn diode at high voltages, known as the Gunn effect¹². The oscillation period of 40 Hz is determined by the switching frequency between the low- and high-current states.

We now discuss the origin of the giant nonlinear resistance. Figure 3 shows the temperature dependence of V_{sample} in sample

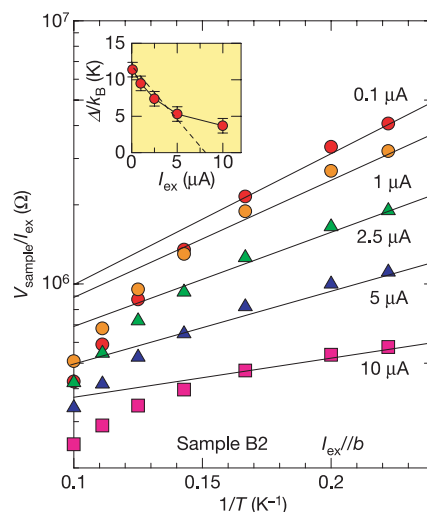


Figure 4 | Nonlinear resistance $V_{\text{sample}}/I_{\text{ex}}$ plotted as a function of $1/T$ for various I_{ex} . The data are taken from Fig. 3. The solid lines are fits to an activated transport using low-temperature data (see text). In the inset the activation energy Δ evaluated from the solid lines is plotted as a function of I_{ex} . Error bars represent the ambiguity in the fits, and show $\pm$ s.d.

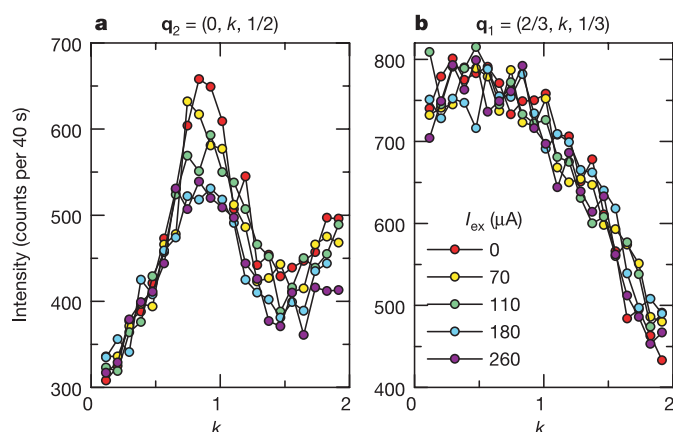


Figure 5 | Diffuse scattering intensities for θ -(BEDT-TTF) $_2$ CsZn(SCN) $_4$ with various external currents applied along the c -axis direction at 12 K. The intensity is plotted as a function of k (along the b^* direction) in units of the reciprocal lattice constant. There are two diffuse streaks at $\mathbf{q}_1 = (2/3, k, 1/3)$ and $\mathbf{q}_2 = (0, k, 1/2)$ corresponding to charge order domains with different patterns. Figures 5a and b show the intensities at $\mathbf{q}_2 + \mathbf{G}_2$ and $\mathbf{q}_1 + \mathbf{G}_1$, respectively, where $\mathbf{G}_1 = (2, 1, 1)$ and $\mathbf{G}_2 = (3, 1, 0)$ are reciprocal lattice vectors chosen suitably for the measurement conditions. Whereas the $\mathbf{q}_2 + \mathbf{G}_2$ streak has a finite interlayer correlation, the $\mathbf{q}_1 + \mathbf{G}_1$ streak does not.

B2. With increasing temperature, the negative derivative resistance rapidly fades away, although the nonlinear resistance survives up to higher temperatures. At low temperatures, charge-ordered domains begin to appear. We expect the charge order to be characterized by a constant energy gap Δ , although the phase may have no long-range order. Then the charge transport should be of activation-type with an activation energy $\sim \Delta$. To include the giant nonlinear conduction effect, we allow Δ to be dependent on I_{ex} , and propose the following phenomenological expression:

$$V_{\text{sample}}/I_{\text{ex}} = R_{\text{CO}} \exp[\Delta_{I_{\text{ex}}}/k_{\text{B}}T] + R_{\text{N}} \quad (1)$$

where R_{CO} and R_{N} are the resistances for the charge-ordered and conducting electrons, respectively. At low temperatures, we can safely neglect R_{N} and the temperature dependence of $\Delta_{I_{\text{ex}}}$. Then $V_{\text{sample}}/I_{\text{ex}}$ plotted on a logarithmic scale as a function of $1/T$ gives $\Delta_{I_{\text{ex}}}$, as shown in Fig. 4. The data show an upward curve, which is consistent with the temperature dependence of the resistivity of the charge-density-wave material $\text{K}_{0.3}\text{MoO}_3$ (ref. 13). This arises from a rapid growth of the gap $\Delta_{I_{\text{ex}}}$ below 20 K, and the temperature-independent $\Delta_{I_{\text{ex}}}$ —that is, $\Delta(I_{\text{ex}}, T) \approx \Delta(I_{\text{ex}}, 0)$ —is only valid for $T \ll 20$ K. We focused on the I_{ex} dependence of $\Delta_{I_{\text{ex}}}$ at 4 K, so we evaluated the magnitude of $\Delta_{I_{\text{ex}}}$ by fitting the data close to 4 K with the solid lines indicated in Fig. 4. Note that the slope, which is equal to $\Delta_{I_{\text{ex}}}/k_{\text{B}}$, decreases with increasing I_{ex} .

The value of Δ extracted from such analysis is plotted as a function of I_{ex} in the inset of Fig. 4. Δ/k_{B} for $I_{\text{ex}} \rightarrow 0$ is 13 K. This value is close to the temperature $T = 20$ K below which the resistivity rapidly increases (a signature of the growth of the charge order), thus validating to some extent our evaluation of Δ . It should be emphasized that Δ decreases roughly linearly with increasing I_{ex} , suggesting that the external current rapidly suppresses the charge order. Accordingly, negative derivative resistance can occur when the suppression of $\exp[\Delta/k_{\text{B}}T]$ is so rapid that $I_{\text{ex}} \exp[\Delta/k_{\text{B}}T]$ becomes a decreasing function of I_{ex} .

We note the following. (1) The negative derivative resistance was observed also in the in-plane (c -axis) direction. Thus, the voltage-current characteristics discussed here are not an interlayer effect. (2)

Essentially identical voltage-current characteristics were observed in θ -(BEDT-TTF) $_2$ CsZn(SCN) $_4$ along the b - and c -axis directions, which are therefore inherent to the Cs salt. Some samples showed a similar a.c. oscillation from d.c. input. (3) The measurements in (1) and (2) above were done in a four-probe configuration with a pulse current source, which indicates that the intrinsic thyristor effects are free from the contact resistance and heating effects.

Finally, to clarify the melting of the charge order by external currents, we show diffraction data under different applied currents. For the experimental requirements, we used θ -(BEDT-TTF) $_2$ CsZn(SCN) $_4$ with the external current applied along the c direction. Figure 5 shows d.c. current dependence of the diffuse scattering intensities at $\mathbf{q}_1 = (2/3, k, 1/3)$ and $\mathbf{q}_2 = (0, k, 1/2)$ at 12 K, which correspond to different charge order domains competing with each other^{5,6}. The intensity was 1,000 times weaker than that of the lattice Bragg reflection, so we used a four-circled diffractometer (wavelength $\lambda = 0.083$ nm) and synchrotron radiation as an X-ray source at BL02B1/SPRING-8, Japan (details in the legend to Fig. 5). As is clearly shown, $\mathbf{q}_2$ loses its intensity with increasing current, while $\mathbf{q}_1$ remains nearly intact. The external current suppresses mainly the charge ordering with the $\mathbf{q}_2$ modulation that is responsible for the insulating behaviour^{2,5,6}, consistent with the current-dependent Δ in Fig. 4.

Received 20 March; accepted 27 July 2005.

1. Streetman, B. G. *Solid State Electronic Devices* 2nd edn, Ch. 11 (Prentice-Hall, Englewood Cliffs, 1980).
2. Mori, H., Tanaka, S. & Mori, T. Systematic study of the electronic state in θ -type BEDT-TTF organic conductors by changing the electronic correlation. *Phys. Rev. B* **57**, 12023–12029 (1998).
3. Seo, H. Charge ordering in organic ET compounds. *J. Phys. Soc. Jpn* **69**, 805–820 (2000).
4. Clay, R. T., Mazumdar, S. & Campbell, D. K. Charge ordering in θ -(BEDT-TTF) $_2$ X materials. *J. Phys. Soc. Jpn* **71**, 1816–1819 (2002).
5. Nogami, Y. *et al.* Structural modulation in θ -(BEDT-TTF) $_2$ CsM'(SCN) $_4$ [M' = Co, Zn]. *Synth. Met.* **103**, 1911 (1999).
6. Watanabe, M., Nogami, Y., Oshima, K., Mori, H. & Tanaka, S. Novel pressure-induced 2k_F CDW state in organic low-dimensional compound θ -(BEDT-TTF) $_2$ CsCo(SCN) $_4$. *J. Phys. Soc. Jpn* **68**, 2654–2663 (1999).
7. Mori, T. Non-stripe charge order in the θ -phase organic conductors. *J. Phys. Soc. Jpn* **72**, 1469–1475 (2003).
8. Nishio, Y. *et al.* Specific heat and metal-insulator transition of (BEDT-TTF) $_2$ MZn(SCN) $_4$ (M = Cs, Rb). *Synth. Met.* **103**, 1905–1906 (1999).
9. Takahashi, T. *et al.* Charge ordering in non-dimerized BEDT-TTF based organic conductors: ¹³C-NMR experiments. *Phys. IV Fr.* **12**(Pr9), 201–204 (2002).
10. Inagaki, K., Terasaki, I., Mori, H. & Mori, T. Large dielectric constant and giant nonlinear conduction in the organic conductor θ -(BEDT-TTF) $_2$ CsZn(SCN) $_4$. *J. Phys. Soc. Jpn* **73**, 3364–3369 (2004).
11. Maeda, A., Notomi, M., Uchinokura, K. & Tanaka, S. Evidence for the existence of the inherent periodicity in the switched state at low temperatures in $\text{K}_{0.3}\text{MoO}_3$. *Phys. Rev. B* **36**, 7709–7711 (1987).
12. Gunn, J. B. Microwave oscillations of current in III–V semiconductors. *Solid State Commun.* **1**, 88–91 (1963).
13. Gruner, G. *Density Waves in Solids* 56–57 (Addison-Wesley, Reading, Massachusetts, 1994).

Acknowledgements We thank K. Inagaki for collaboration, and A. Maeda for technical advice for nonlinear-conduction and noise measurements. We also thank S. Tasaki, S. Kurihara, M. Abdel-Jawad, and N. E. Hussey for discussions. This work was partially supported by MEXT, the Grant-in-Aid for Scientific Research, and by the 21st Century COE Program at Waseda University.

Author Contributions F.S. did the electrical measurement, I.T. did the project planning and analysis, H.M. and T.M. did the sample preparation and chemical characterization, and M.W., N.I., Y.N. and Y.N. did the diffraction in electric fields.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to I.T. (terra@waseda.jp).

Optical isotropy and iridescence in a smectic 'blue phase'

Jun Yamamoto^{1,2†}, Isa Nishiyama¹, Miyoshi Inoue¹ & Hiroshi Yokoyama^{1,2}

When liquid crystal molecules are chiral, the twisted structure competes with spatially uniform liquid crystalline orders, resulting in a variety of modulated liquid crystal phases, such as the cholesteric blue phase¹, twist grain boundary^{2–4} and smectic blue phases⁵. Here we report a liquid crystal smectic blue phase (SmBP_{iso}), formed from a two-component mixture containing a chiral monomer and a 'twin' containing two repeat units of the first molecule connected by a linear hydrocarbon spacer. The phase exhibits the simultaneous presence of finite local-order parameters of helices and smectic layers, without any discontinuity on a mesoscopic length scale. The anomalous softening of elasticity due to a strong reduction in entropy caused by mixing the monomer and the twin permits the seamless coexistence of these two competing liquid crystal orders. The new phase spontaneously exhibits an optically isotropic but uniformly iridescent colour and automatically acquires spherical symmetry, so that the associated photonic band gap^{6–9} maintains the same symmetry despite the local liquid crystalline order. We expect a range of unusual optical transmission properties based on this three-dimensional isotropic structure, and complete tunability due to the intrinsic softness and responsiveness of the liquid crystalline order against external fields.

We have designed a two-component mixture that consists of a chiral monomer (3B1M7) and its twin (BMHBOP-6)¹⁰. The twin consists of two equivalent monomer units connected by a linear hydrocarbon spacer, as shown in Fig. 1. We termed it a 'commensurate' mixture because the length of the twin is exactly twice that of the monomer. We found that mixing the twin into monomer is apparently equivalent to the introduction of connecting chains between two monomers, which freeze the motion of the monomers. There are two concentrations— ϕ_m and ϕ_c —that are equivalent to the concentrations of the monomer and (*R,R*) enantiomer in the twin component, respectively.

We note that six modulated phases appear in the mixture, as shown in the phase diagram (Fig. 1), but only the isotropic liquid (Iso) and smectic-A (*S_A*) phases appear in a pure monomer. We identified three different SmBPs—SmBP_{X1}, SmBP_{X2} and SmBP_{X3}. They appear to be similar to the previously reported phases SmBP₁, SmBP₂ and SmBP₃, respectively⁵. We also identified a fourth smectic blue phase, SmBP_{iso}. A characteristic feature of SmBP_{iso} is that it exhibits a completely uniform colour under a polarizing microscope without any type of macroscopic pattern or texture. Moreover, the colour does not change when the sample is rotated in any direction. This proves that SmBP_{iso} has a completely isotropic optical property. It should be noted that SmBP_{iso} can be detected only within a very narrow range of ϕ_m from 0.65 to 0.75. SmBP_{X1} and SmBP_{X2} can be distinguished clearly from SmBP_{iso} by their characteristic platelet textures, as shown in Fig. 2a and b, respectively.

During the cooling process, SmBP_{iso} appeared below 127 °C and was identified by the appearance of a uniform iridescent colour, as shown in Fig. 2b. The wavelength of the colour increased with a decrease in the temperature, which is a unique feature independent of both ϕ_m and ϕ_c . The platelet texture of SmBP_{X2} appeared below 124 °C. During the reversible heating process, the platelet texture spontaneously melted above 124 °C, which is the temperature at which the uniform colour disappeared during the cooling process, into the uniformly coloured texture of SmBP_{iso}. The temperature dependence of the colour of SmBP_{iso} was identical for both the cooling and heating processes, and no hysteresis was observed.

Thus, Fig. 2b provides two crucial experimental results. First, all phases (SmBP_{X3}, SmBP_{iso} and SmBP_{X2}) can be proved to be thermal equilibrium phases because both phase transitions—SmBP_{X3} to SmBP_{iso} and SmBP_{iso} to SmBP_{X2}—were confirmed to be thermodynamically reversible. Second, the isotropic nature of SmBP_{iso} arises from an inherent feature of its internal structure, and it is not related to the existence of a polycrystalline structure during the dynamical process of phase transition. The phase transition (and colour change) is quite smooth and rapid without any time delay during the nucleation or crystal-growth processes.

Spectroscopy experiments with polarized visible light revealed that both the optical rotatory power and transmitted light intensity of the linearly polarized light in SmBP_{iso} exhibited a large wavelength

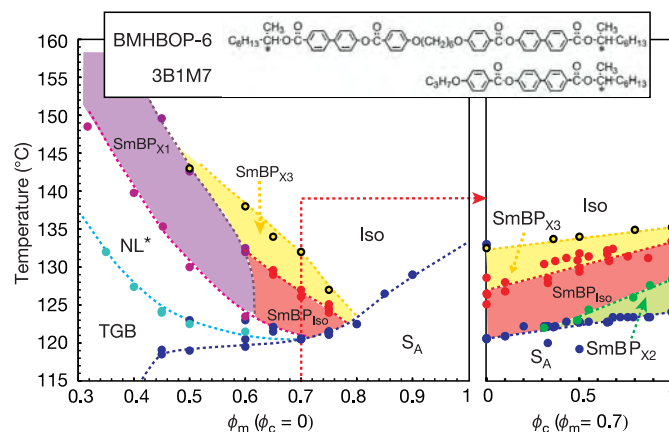


Figure 1 | Phase diagram for the twin/monomer mixture. Left, the monomer concentration (ϕ_m) dependence; right, the optical purity (ϕ_c) dependence of the twin by fixing $\phi_m = 0.7$. The TGB and NL* phases can easily be identified under a polarizing microscope because of their large optical anisotropy and characteristic patterns. The inset shows the molecular formulas of the twin (BMHBOP-6) and monomer (3B1M7).

¹ERATO Yokoyama Nano-structured Liquid Crystal Project, JST, 5-9-9 Tokodai, Tsukuba 300-2635, Japan. ²Nanotechnology Research Institute, AIST, 1-1-1 Umezono, Tsukuba 305-8568, Japan. [†]Present address: Department of Physics, Graduate School of Science, Kyoto University, Kitashirakawa, Sakyo, Kyoto 606-8502, Japan.

dependence, as shown in Fig. 3a. The spectrum did not change when the sample was rotated in any direction, which again confirmed the isotropic optical property. From the following results, we can conclude that the colour originates from the spatial distribution of a helix, which is similar to that of the cholesteric (Ch) phase. (1) In the short-wavelength region, the optical rotatory power diverges and an extremely small amount of light is transmitted. The short-wavelength region is similar to the pitch band of selective reflection, which is commonly observed in the Ch phase. (2) The short-wavelength cut-off is independent of the sample thickness d (100 μm –1 mm), whereas the reflectivity increases only with an increase in d . (3) The transmittance of circularly polarized light is completely dependent on the optical handedness of the mixture.

Thus, these results confirm the existence of the helix and its isotropic distribution in space. Moreover, the transmittance of light at a greater wavelength, far from the band edge, is close to unity even in the case of an extremely thick sample (~ 1 mm), which indicates the absence of any domain structure with a length scale greater than the helix pitch. During the X-ray experiment, only one broad scattering peak was observed in the small angle region. This is a direct confirmation that all six modulated phases—twist grain boundary (TGB), chiral line nematic (NL^*), SmBP_{X1} , SmBP_{X2} , SmBP_{X3} and SmBP_{Iso} —have a finite smectic order with the exception of the order parameter difference. As shown in Fig. 3b, the peak intensity begins to grow continuously from SmBP_{X3} to the S_A phase. The characteristic length derived from the peak position is smoothly connected to the layer repeat distance of S_A in all the other phases. Thus, we can identify only the Iso to SmBP_{X3} phase transition in the X-ray results. In contrast, all the modulated phases are clearly distinguished by their macroscopic viscoelastic properties¹¹, as shown in Fig. 4a. In SmBP_{X3} , viscoelastic relaxation appears and a frequency-dependent real part is observed; however, the fluidity

remains constant. We note that a frequency-independent finite real part appears in SmBP_{Iso} . The static elasticity of SmBP_{Iso} is approximately one order smaller than the layer compression modulus of S_A .

The most important and unusual characteristics of SmBP_{Iso} are that it spontaneously exhibits an optically isotropic but uniformly iridescent colour. X-ray and visco-elastic measurements clearly demonstrate that all SmBPs have a smectic order, and the simultaneous existence of the helices was confirmed from the optical rotatory power. Thus, two different types of liquid crystalline orders coexist in space with frustration in all SmBPs . A universal feature is that spontaneous symmetry breaking leads to a liquid crystalline order, which results in the appearance of spatial macroscopic anisotropy. However, the spherical symmetry only can be restored spontaneously by randomizing the regular spatial distribution of both the liquid crystalline orders on a larger scale in SmBP_{Iso} . The texture of SmBP_{Iso} is clearly smooth and homogeneous without any discontinuity. If a system has a true isotropic symmetry, spatial uniformity is automatically realized without any internal domain

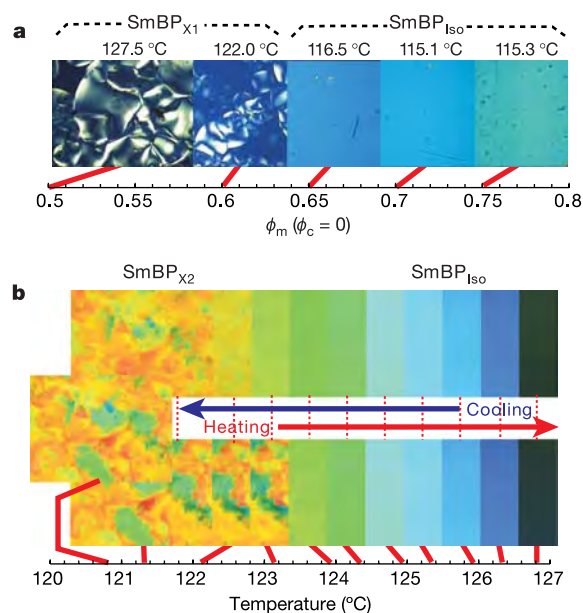


Figure 2 | Polarizing microscope photographs in SmBPs. **a**, ϕ_m dependence of the patterns of SmBP_{X1} and SmBP_{Iso} by fixing the optical purity of the twin ($\phi_c = 0$). Red bars indicate the concentration of the mixture. The characteristic platelet texture of the SmBP_{X1} phase was observed below $\phi_m = 0.6$. The SmBP_{Iso} texture was observed to be of uniform colour from $\phi_m = 0.65$ – 0.75 . **b**, Temperature dependence of the patterns during both the cooling and heating processes of the SmBP_{Iso} to SmBP_{X2} phase transition. Red bars indicate the actual temperature of the sample. A sample ($\phi_m = 0.7$ and $\phi_c = 1$) was introduced in a 1-mm-thick quartz cell. The thermo-reversible SmBP_{Iso} to SmBP_{X2} phase transition is clearly identified at approximately 124 °C.

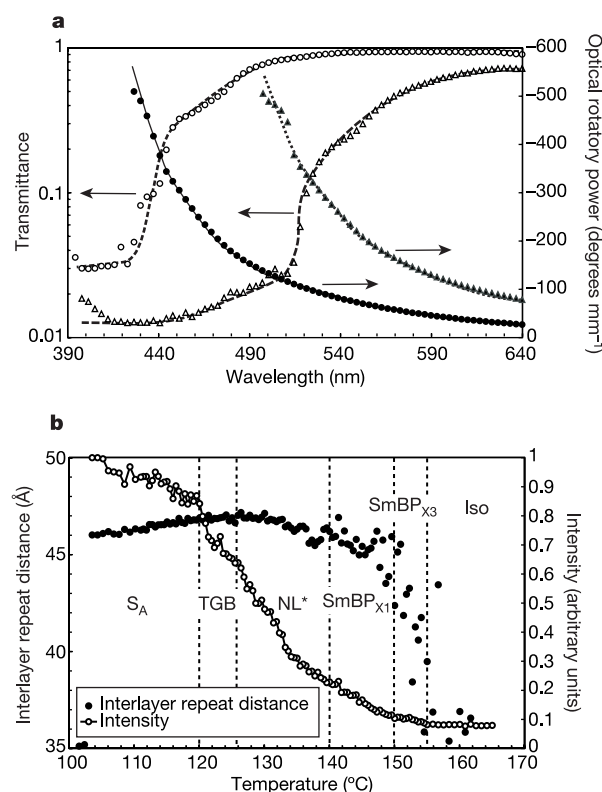


Figure 3 | Macroscopic and microscopic structural analysis in SmBPs. **a**, Transmittance (open circles) and optical rotatory power (closed circles) of SmBP_{Iso} ($T = 124.7$ °C) deduced from the analyser angle dependence of the spectrum of the polarized visible light ($\phi_m = 0.7$, $\phi_c = 1$). The transmittance decreases by less than 2% in the short-wavelength region below 430 nm whereas there is complete transparency above 600 nm. The optical rotatory power also diverges in the short-wavelength region, which is similar to the pitch band of selective reflection in the cholesteric phase. It is remarkable that the edge of the pitch band continuously shifted to 530 nm for the supercooled SmBP_{Iso} ($T = 120.2$ °C), as indicated by the triangular symbols (open and closed symbols represent the transmittance and optical rotatory power, respectively). **b**, Temperature dependence of the smectic order measured by X-ray scattering obtained from a sample ($\phi_m = 0.45$, $\phi_c = 0$). Only one scattering peak was detected in the small-angle region below 156 °C, which confirms that with the exception of the Iso phase, all the phases have a finite local smectic order. The intensity of the scattering peak increases continuously from SmBP_{X3} to S_A , as indicated by the open circles. The characteristic length is smoothly connected to the layer repeat distance of S_A , as indicated by the filled circles.

structures. This implies that the symmetry of the novel SmBP_{Iso} is intrinsically spherical.

To establish the details of the internal structure of SmBP_{Iso} , we first note that both the layer compression modulus B and Frank elastic constant K_1 in the S_A phase are minimized at approximately $\phi_m = 0.7$, as shown in Fig. 4b. At this value of ϕ_m , the value of B is 20 times smaller, while that of K_1 is almost two orders of magnitude smaller than their values in the pure-monomer system. By comparing Fig. 1 and Fig. 4b, we observe that SmBP_{Iso} was stabilized only within a very narrow concentration range in which a drastic softening of B and K_1 occurs. SmBP_{Iso} certainly results from the anomalous

softening of the smectic layer. Using this information, we propose a rough sketch of the molecular model, as shown in inset (1) of Fig. 4a. There are two energy minima with respect to the relative position of the monomer and twin. First, the centre of gravity of all the molecules must lie on the mid-plane of the smectic layer, owing to the requirement of fluctuation motion. This is completely equivalent to our concept of molecular cooling of the twin molecules. Second, another local energy minimum is due to the local molecular interactions. We emphasize that the commensurability of our mixture could enhance the effect of the lateral interaction. Thus, we can assume that a continuous slip of the smectic layer can easily be generated by the localization of the twin without decreasing the defect core energy, as shown in inset (2) of Fig. 4a.

As is evident in the freeze-fracture replica and frozen replica of the free air interface of TEM photographs of SmBP_{Iso} (see Supplementary Figs 1, 2 and 3), we found a random and continuous interconnected sponge-like structure in SmBP_{Iso} . The characteristic size of this structure exhibits a large dispersion around the visible light wavelength. A multiple stack of the smectic layers in the interconnected structure is also shown in Supplementary Figs 2 and 3. A multi-lamellar interconnected cubic structure was recently proposed theoretically¹²; however, the proposed structure has spatial regularity because it is a model of the optically cubic crystals of SmBP_{X1} and SmBP_{X2} . In fact, these SmBP s exhibit angle-dependent selective reflection owing to the regular spatial arrangement of the helices, which appears to be identical to the double helix structure in ChBP s. Therefore, the lattice of the multi-lamellar interconnected structure must coexist and be frustrated with the lattice of the double helices.

Thus we can consider that the frustration induces the melting of the lattice of the multi-lamellar interconnected structure in a manner related to the liquid crystal microemulsion¹³. A random sponge-like distribution is then produced, which is similar to the well-known cubic-sponge phase transition of the unilamellar interconnected structure in the lyotropic system. We conclude that SmBP_{Iso} is the multi-lamellar sponge phase. Furthermore, the SmBP_{X2} to SmBP_{Iso} transition can be attributed to the newly discovered multi-lamellar cubic-sponge phase transition. Neither of the liquid crystalline orders, that is the smectic layer or the helix, disappear during these phase transitions. The multi-lamellar interconnected units are also maintained in both phases, and only their spatial regularity is lost in SmBP_{Iso} .

Theoretical calculations of the driving force of the multi-lamellar interconnected structure¹² considered only the gaussian curvature energy without chiral interaction. However, our system is extremely sensitive to optical purity. We suggest that the chirality promotes the generation of a negative gaussian curvature of the multilayer owing to the coupling between the chiral interaction and layer deformation. Another theoretical work predicted that the bending of the layer should be induced by molecular chirality^{14,15}. The behaviour of the B_4 phase in the original banana molecules¹⁶ is similar to that of SmBP s. However, a clear mosaic texture and weak birefringence are always observed in the B_4 phase under a polarizing microscope¹⁷. These characteristics completely contradict the spherical symmetry of SmBP_{Iso} . On the other hand, the banana and twin (dimmer) molecules have a similar molecular structure. As compared to the bilayer membrane formed by the molecules with strong steric incompatibility between the head and tail, it is reasonable that the smectic layer, which consists of the banana or twin molecules, should prefer the saddle-splay deformation of the layers, which is a common fundamental unit for all SmBP s, not only for SmBP_{Iso} .

In summary, we have experimentally found a novel smectic phase SmBP_{Iso} , which is optically isotropic, has true spherical symmetry and exhibits iridescent colour. However, the distribution of the helix in the multi-lamellar interconnected structure is still unclear. It is highly advantageous that the characteristic length of the system can be completely tuned in the visible light wavelength by external factors such as temperature and light irradiation. This is because the

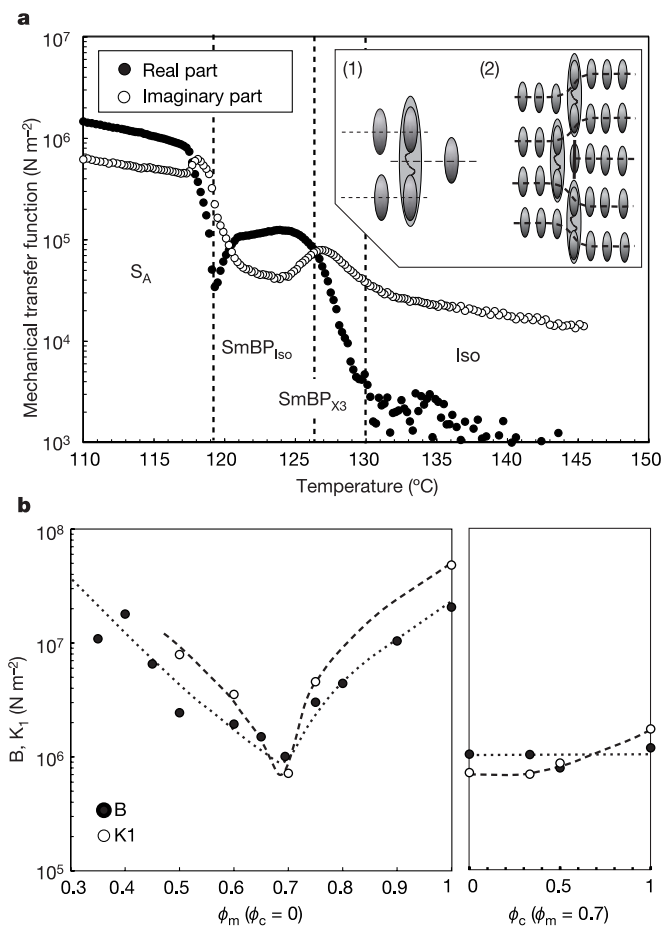


Figure 4 | Mechanical properties in SmBP s. **a**, Temperature dependence of the real (filled circles) and imaginary (open circles) parts of the complex mechanical transfer function of the sample ($\phi_m = 0.7$, $\phi_c = 0$) at 200 Hz. Both the SmBP_{X3} and SmBP_{Iso} phases can be observed. It is evident that the finite static elastic part observed in the SmBP_{Iso} phase is one order of magnitude smaller than the layer compression modulus of the S_A phase. The small value of the real part in the SmBP_{X3} phase is due to the viscoelastic relaxation. The insets show a schematic representation of the molecular models: (1) There are two energy minima with respect to the relative position of the monomer and twin. The position of the dotted line corresponds to the stable position based on the entropic origin. The centre of gravity of the twin and monomer, namely the centre of the fluctuation motion, should be settled on the centre of the layers. On the other hand, the position of the dotted line corresponds to the lateral interaction of the local intermolecular interaction. The chemical contrast could be enhanced by the commensurability of the system. Each position was shifted merely by half the monomer length. (2) A defect can be induced easily in the layer by the localization of the twin molecules. **b**, ϕ_m and ϕ_c dependencies of the layer compression modulus B (red circles) and Frank elastic constant K_1 (blue circles) in the S_A phase. Our viscoelastic measurement can simultaneously determine B and K_1 by detecting the dynamic change in B immediately after the application of a large step-like dilative strain.

photonic bandgap is formed by a self-organized hierarchical structure driven by the frustration in the soft liquid crystalline order. Until now, such tunability has not been achieved even by using photonic crystal materials that were fabricated well. In fact, a similar isotropic medium has been designed by modelling the complete randomness in polycrystalline domains. An artificial design for a self-organized hierarchical structure must be compatible with bottom-up 'nanotechnology' to use the frustration induced by impurities in the liquid crystalline orders¹³. These long-range elastic interactions can stabilize huge structures such as biological systems in which short-range molecular interactions become completely irrelevant.

Received 24 April; accepted 30 June 2005.

1. Ketzerow, H. S. & Bahr, C. (eds) *Chirality in Liquid Crystals* 186–222 (Springer, New York, 2002).
2. Renn, S. R. & Lubensky, T. C. Abrikosov dislocation lattice in a model of the cholesteric-to-smectic A transition. *Phys. Rev. A* **38**, 2132–2147 (1988).
3. Goodby, J. W. *et al.* Characterization of a new helical smectic liquid crystal. *Nature* **337**, 449–452 (1989).
4. Dierking, I. & Lagerwall, S. T. A review of textures of TGB* phase under different anchoring conditions. *Liq. Cryst.* **26**, 83–95 (1999).
5. Pansu, B., Grelet, E., Li, M. H. & Nguyen, H. T. Hexagonal symmetry for smectic blue phases. *Phys. Rev. E* **62**, 658–665 (2000).
6. John, S. Electromagnetic absorption in a disordered medium near photon mobility edge. *Phys. Rev. Lett.* **53**, 2169–2172 (1984).
7. Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
8. John, S. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
9. Busch, K. & John, S. Liquid-crystal photonic-band-gap materials: the tunable electromagnetic vacuum. *Phys. Rev. Lett.* **83**, 967–970 (1999).
10. Nishiyama, I., Yamamoto, J., Goodby, J. W. & Yokoyama, H. Symmetric chiral liquid crystalline twin exhibiting stable ferroelectric and antiferroelectric phases and a chirality-induced isotropic-isotropic liquid transition. *J. Mater. Chem.* **11**, 2690–2693 (2001).
11. Yamamoto, J. & Okano, K. Anomalous hydrodynamic behaviors of smectic liquid crystals at low frequencies. *Jpn. J. Appl. Phys.* **30**, 754–763 (1991).
12. DiDonna, B. A. & Kamien, R. D. Smectic phases with cubic symmetry: the splay analog of the blue phase. *Phys. Rev. Lett.* **89**, 215504 (2002).
13. Yamamoto, J. & Tanaka, H. Transparent nematic phase in liquid-crystal based microemulsion. *Nature* **409**, 321–325 (2001).
14. MacKintosh, F. C. & Lubensky, T. C. Orientational order, topology, and vesicle shapes. *Phys. Rev. Lett.* **67**, 1169–1173 (1991).
15. Zhong-can, O. & Ji-xing, L. Helical structure of tilted chiral lipid bilayers viewed as cholesteric liquid crystals. *Phys. Rev. Lett.* **65**, 1679–1682 (1990).
16. Niori, T., Sekine, T., Watanabe, J., Furukawa, T. & Takezoe, H. Distinct ferroelectric smectic liquid crystals consisting of banana shaped achiral molecules. *J. Mater. Chem.* **6**, 1231–1233 (1996).
17. Thisayukta, J., Takezoe, H. & Watanabe, J. Study on helical structure of the B₄ phase formed from achiral banana-shaped molecules. *Jpn. J. Appl. Phys.* **40**, 3277–3287 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Hough and N. A. Clark for suggesting methods of preparing the frozen replica of the TEM observation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.Y. (jun@junyamamoto.jp).

Europe-wide reduction in primary productivity caused by the heat and drought in 2003

Ph. Ciais¹, M. Reichstein^{2,3}, N. Viovy¹, A. Granier⁴, J. Ogée⁵, V. Allard⁶, M. Aubinet⁷, N. Buchmann⁸, Chr. Bernhofer⁹, A. Carrara¹⁰, F. Chevallier¹, N. De Noblet¹, A. D. Friend¹, P. Friedlingstein¹, T. Grünwald⁹, B. Heinesch⁷, P. Keronen¹¹, A. Knohl^{12,13}, G. Krinner¹⁴, D. Loustau⁵, G. Manca^{2†}, G. Matteucci^{15†}, F. Miglietta¹⁶, J. M. Ourcival¹⁷, D. Papale², K. Pilegaard¹⁸, S. Rambal¹⁷, G. Seufert¹⁵, J. F. Soussana⁶, M. J. Sanz¹⁰, E. D. Schulze¹², T. Vesala¹¹ & R. Valentini²

Future climate warming is expected to enhance plant growth in temperate ecosystems and to increase carbon sequestration^{1,2}. But although severe regional heatwaves may become more frequent in a changing climate^{3,4}, their impact on terrestrial carbon cycling is unclear. Here we report measurements of ecosystem carbon dioxide fluxes, remotely sensed radiation absorbed by plants, and country-level crop yields taken during the European heatwave in 2003. We use a terrestrial biosphere simulation model⁵ to assess continental-scale changes in primary productivity during 2003, and their consequences for the net carbon balance. We estimate a 30 per cent reduction in gross primary productivity over Europe, which resulted in a strong anomalous net source of carbon dioxide (0.5 Pg C yr⁻¹) to the atmosphere and reversed the effect of four years of net ecosystem carbon sequestration⁶. Our results suggest that productivity reduction in eastern and western Europe can be explained by rainfall deficit and extreme summer heat, respectively. We also find that ecosystem respiration decreased together with gross primary productivity, rather than accelerating with the temperature rise. Model results, corroborated by historical records of crop yields, suggest that such a reduction in Europe's primary productivity is unprecedented during the last century. An increase in future drought events could turn temperate ecosystems into carbon sources, contributing to positive carbon-climate feedbacks already anticipated in the tropics and at high latitudes^{1,2}.

Europe experienced a particularly extreme climate anomaly during 2003, with July temperatures up to 6 °C above long-term means, and annual precipitation deficits up to 300 mm yr⁻¹, 50% below the average. The presence of an extensive network of instrumentation for the monitoring of ecosystem fluxes at this time, with continuous records of CO₂, water, and energy fluxes^{7,8}, helped us to assess the impact of such an extreme event on the continental-scale carbon balance. We analysed CO₂ fluxes from 14 forest sites and one grassland site for 2002–2003 (Table 1). Flux records are generally not yet long enough to provide long-term average references, and therefore we used 2002 as a reference (see Methods). Hourly fluxes of photosynthesis (gross primary productivity, GPP) and total

ecosystem respiration (TER) are separated from CO₂ net fluxes (net ecosystem exchange, NEE), using the same method for each site⁹.

Given inter-site differences in drought duration and intensity, soil characteristics, vegetation state, and species-specific responses to climate variation, one would not expect a uniform response of GPP to the abnormal conditions of 2003. Yet, Table 1 (also Supplementary Fig. S1) clearly shows that nearly all sites experienced a significant GPP reduction in 2003. The GPP drop coincides with reduced evapotranspiration and soil drying due to the rainfall deficit (Supplementary Fig. S1). Generally, below a threshold of ~0.4 in relative extractable water, water stress occurs, causing both GPP and transpiration to decrease in response to stomatal closure^{10,11}. Particularly large reductions in GPP were found in temperate deciduous beech and northern Mediterranean forests (Hesse, Hainich, Roccarespanpani, San Rossore), together with reductions in canopy conductance (in 2003, conductance reached only 15% of its 2002 value at Hesse). These productive temperate and Mediterranean forests were greatly affected by extreme drought and/or heat (Fig. 1). Moreover, the GPP did not entirely recover from the summer stress during the remainder of the growing season (Fig. 1). Southern evergreen forest sites El Saler (pine) and Puéchabon (oak) also experienced a reduction in GPP during the heatwave (Table 1). Southern forest sites with an herbaceous layer that normally dries in summer were less affected by the heatwave and drought (Pianosa).

Warmer temperatures are expected to increase both microbial and plant respiration. However, during the 2003 summer, TER fell in parallel with GPP at most sites (Table 1, Supplementary Fig. S1). A reduction in both plant respiration (due to diminished substrates) and microbial soil respiration (also due to drought) can explain such parallel TER and GPP responses. Possible artefacts of the method used to separate gross fluxes⁹ can largely be ruled out because of the close correlation between the independent quantities TER and mid-day NEE, the latter dominated by GPP (Fig. 1). Overall, those forests where GPP strongly decreased also had reduced net carbon uptake (NEE), with Hesse and Tharandt even temporarily becoming net

¹Laboratoire des Sciences du Climat et de l'Environnement, LSCE, F-91191, Gif sur Yvette, France. ²Department of Forest Environment and Resources, DISAFRI, University of Tuscia, I-01100 Viterbo, Italy. ³Potsdam Institute for Climate Impact Research, Telegrafenberg C4, D-14473 Potsdam, Germany. ⁴Ecologie et Ecophysiologie Forestières, Centre de Nancy, F-54280 Champenoux, France. ⁵Functional Ecology and Environmental Physics, Ephyse, INRA, F-33612 Villenave d'Ornon, France. ⁶Grassland Ecosystem Research, INRA, 234 Avenue du Brézet, Clermont-Ferrand, F-63039, France. ⁷Institute of Plant Sciences, Universitätsstrasse 2, ETH Zentrum LFW C56, CH-8092 Zuerich, Switzerland. ⁸Faculté des Sciences Agronomiques, Avenue de la Faculté d'Agronomie 8, B-5030 Gembloux, Belgium. ⁹Department of Meteorology, Institute of Hydrology and Meteorology, Technische Universität Dresden, D-01062 Dresden, Germany. ¹⁰Fundación CEAM C/ Charles Darwin, Parque Tecnológico de Paterna, E-46980 Valencia, Spain. ¹¹Department of Physical Sciences, University of Helsinki PO-Box 64, Finland. ¹²Max-Planck Institute for Biogeochemistry, Postfach 10 01 64, D-07701 Jena, Germany. ¹³Department of Environmental Science Policy and Management, Ecosystem Science Division, University of California, Berkeley, California 94720, USA. ¹⁴Laboratoire de Glaciologie et Géophysique de l'Environnement, 54 rue Molière, F-38402, Saint Martin d'Hères, France. ¹⁵Institute for Environment and Sustainability, Joint Research Center European Commission, TP 280, I-21020 Ispra, Italy. ¹⁶BIMET-CNR, Ple delle Cascine, 18, I-50144 Firenze, Italy. ¹⁷Dream CEEF-CNRS, 1919 route de Mende, F-34293 Montpellier, France. ¹⁸Biosystems Department, Risø National Laboratory, 4000 Roskilde, Denmark. †Present addresses: Centro di Ecologia Alpina, Viote del Monte Bondone, I-38040 Trento, Italy (G. Manca); ISAFOM-CNR, Via Cavour 4-6, I-87036, Rende, Italy (G. Matteucci).

Table 1 | Changes in climate and ecosystem CO₂ fluxes between 2002 and 2003 at eddy covariance sites

Site Code	Name	Vegetation, main genus	Country	Latitude	Longitude	July–September						Annual			
						ΔT	ΔP	ΔGPP	ΔTER	ΔNEP	s.e.*	ΔGPP	ΔTER	ΔNEP	s.e.*
SA	El Saler	ENF, pine	SP	39.28	0.33	1.7	−34	−33	−63	30	7	−94	−460	366	43
CP	Castelporziano	EBF, oak	IT	41.71	12.38	3.5	−42	−47	−21	−26	11	−17	16	−33	86
RO	Roccarespampani	DBF, oak	IT	42.39	11.92	2.3	−118	−117	−89	−28	15	−130	−287	158	95
SR	San Rossore	ENF, pine	IT	43.71	17.28	1.8	−120	−87	−47	−40	11	−344	−292	−51	25
BX	Bray	ENF, pine	FR	44.72	−0.77	2.9	−4	29	21	8	14	180	−114	294	77
LA	Laqueuille	GRA, grass	FR	45.64	2.75	3.5	−15	−25	−19	−6	4	−†	−†	−†	−†
PI	Pianosa	OSH, juniper	IT	42.58	10.07	3.2	−69	5	9	−4	16	−†	−†	−†	−†
PU	Puéchabon	EBF, oak	FR	43.73	3.58	2.2	+3	−52	−24	−28	6	−206	−91	−115	32
HE	Hesse	DBF, beech	FR	48.67	7.08	2.0	−53	−115	−42	−73	9	−291	−187	−104	46
VI	Vielsalm	MF, beech and fir	BE	50.30	6.00	1.4	−18	−20	−37	+17	15	−95	−203	108	75
TH	Tharandt	ENF, spruce	GE	50.95	13.57	1.0	−121	−41	−10	−31	7	−208	−53	−155	53
HA	Hainich	DBF, beech	GE	51.07	10.5	1.8	−30	−82	−25	−57	6	−195	−125	−70	48
SO	Soroe	DBF, beech	DK	55.48	11.63	0.3	−57	−15	−14	−1	7	−158	−183	26	66
HY	Hyttiälä	ENF, pine	FI	61.85	24.28	−0.1	−5	−3	10	−13	5	−52	48	−100	30

The more positive GPP and NEE are, the larger is the carbon uptake from the atmosphere. For the July to September period, units are in $\text{g C m}^{-2} \text{month}^{-1}$, for the annual period in $\text{g C m}^{-2} \text{year}^{-1}$. The more positive TER is, the larger is the carbon release. Changes in climate and CO₂ fluxes are reported for the averaging period of July to September, and for the whole year. The vegetation is coded according to the IGBP classification: ENF, evergreen needle-leaf forest; EBF, evergreen broad-leaf forest; DBF, deciduous broad-leaf forest; GRA, grassland; MF, mixed forest; OSH, open shrubland.

* Standard error (s.e.) estimates for the changes ΔGPP and ΔTER (see Methods for details).

† Non-reported values with more than 20% non-reliably filled gaps in 2002 or 2003.

CO₂ sources to the atmosphere in August. Despite the general reduction of carbon sink of most European sites, it seems that several Mediterranean sites showed a smaller decrease in NEE, largely dominated by less respiration in 2003. Although the eddy-covariance data provide an unambiguous evidence for a spectacular reduction in GPP and in NEE in European forests in 2003 (Table 1) it is still too early to assess the impacts on the long-term carbon balance. Tree damage, changes in litterfall rates, and changes in the pool sizes of carbon reserves will have consequences beyond the duration of the extreme climate event¹². Analysis of the responses to these disturbances should be carried out over the next few years.

For croplands, we analysed harvest data from country-level statistics¹³. In each country, harvest was converted to crop net primary productivity (NPP), using allometric relationships¹⁴. Differences in crop NPP in 2003 versus 1998–2002 reflect the combined response of cultivated plants to climate stress and possible management adaptations to it (for example, increased irrigation). Nevertheless, the harvest data show a pronounced NPP decrease in 2003 (Supplementary Fig. S2) in those agricultural regions affected by heat (northern Italy, France) and by drought (Ukraine, Romania). A record NPP drop of 36% occurred in Italy for maize, a cereal grown in the Po valley where extremely high temperatures prevailed (see Fig. 2 and http://www.esa.int/export/esaEO/SEMZP6YO4HD_index_0.html). Winter crops (wheat) had nearly terminated their growth by the time of the heatwave and therefore suffered less NPP reduction than summer crops (corn) undergoing maximum foliar development (Supplementary Fig. S2). Mediterranean countries normally experience dry and hot summers, and therefore both irrigation and cultivation of drought-tolerant species reduced the impact of the climate conditions in these regions.

We next estimated the Europe-wide impacts of the anomalous 2003 climate on productivity, using a process-based ecosystem model⁵, which calculates phenology, carbon pool dynamics, and hourly fluxes of CO₂, energy and water vapour (see Methods). First, we ran the model at each eddy-covariance site (Supplementary Figs S1 and S3) to verify its ability to reproduce the observed GPP and TER anomalies (data-model average correlation $R^2 = 0.6$ for 2003 versus 2002 anomalies). Second, we simulated the Europe-wide changes (Fig. 2) in carbon fluxes from 1900 to 2003 using reconstructed climate and weather analyses^{15,16}.

We first verified the model-predicted changes in leaf area index (foliar surface per m²) against satellite observations of the fraction of absorbed photosynthetically active radiation (FAPAR) derived from the EOS-Terra-MODIS instrument¹⁷. Simulated and satellite-derived

patterns of FAPAR changes in 2003 versus 2000–2002 show good agreement over western Europe (Fig. 2e, f) with maximum reduction in satellite FAPAR over France (−8%) and Northern Italy (−15%). Over eastern Europe however, we simulated a larger FAPAR reduction than in the satellite observations.

Next, we mapped productivity changes. Although 2003 is not the driest year on record, the impacts of drought on GPP and NPP was amplified by high summer temperatures and soil water deficits carried over from the previous spring. For European forests, we calculate a mean reduction in NPP of $16 \text{ g C m}^{-2} \text{month}^{-1}$ in the summer of 2003 compared to 1998–2002 (Fig. 2c), corresponding to

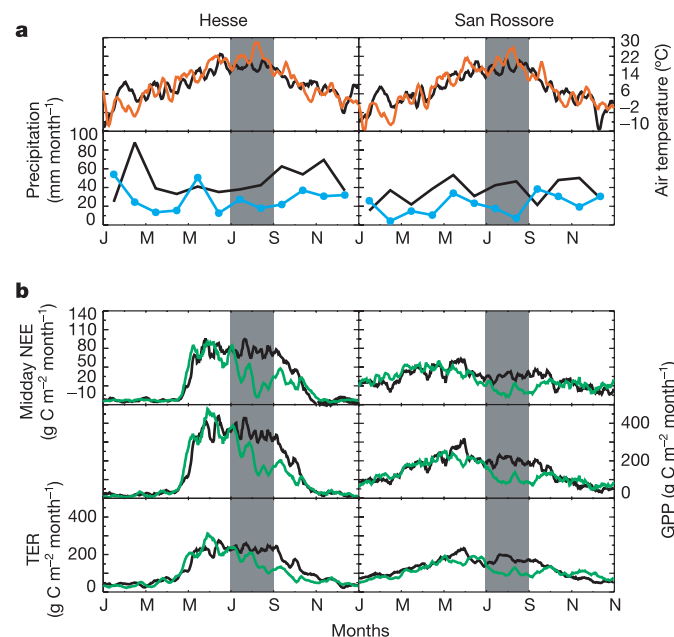


Figure 1 | Observed climate and ecosystem CO₂ fluxes during 2002 and 2003 at two forest sites. The two sites are a temperate deciduous beech forest in France (Hesse) and a southern evergreen pine forest site (San Rossore) in northern Italy. **a**, Climate fields. **b**, Ecosystem CO₂ fluxes. A five-day running average was applied to the original half-hourly flux and temperature data to remove diurnal variations. Precipitation values are monthly averages. Data for 2002 are in black and for 2003 in colour. The July to August period is shaded in grey.

a GPP reduction of $28 \text{ g C m}^{-2} \text{ month}^{-1}$. This finding is consistent with eddy-covariance observations (Table 1). Moreover, as observed at the eddy-covariance sites, a modelled soil water deficit extending from spring to autumn resulted in a significantly lower annual average NPP during 2003 ($541 \text{ g C m}^{-2} \text{ yr}^{-1}$) compared to 1998–2002 ($644 \text{ g C m}^{-2} \text{ yr}^{-1}$). Although our model does not include crop-specific parameterizations, the simulated NPP decrease at crop locations compares well with the crop harvest data (Supplementary Fig. S2). Overall, the NPP reduction in 2003 (Fig. 2c, d) follows the pattern of drought in Central and Eastern Europe, and of extreme summer heat in Western Europe. Accordingly, we find the largest NPP reduction in the Ukraine and Romania (-20%), France (-17%) and Italy (-12%), whereas NPP even increases in southern Sweden in response to moderate warming and no marked water deficits. Overall, the simulated GPP anomalies at the continental scale in 2003 correlate better with rainfall changes than with summer air temperature changes (Supplementary Fig. S4), indicating the dominant role of water limitations.

Modelled respiration (TER) fell over Europe in 2003 by $77 \text{ g C m}^{-2} \text{ yr}^{-1}$, tailing off with a larger GPP reduction of $195 \text{ g C m}^{-2} \text{ yr}^{-1}$, as observed at the eddy-covariance sites. In the

model, the response of TER can be attributed to reduced plant respiration (less assimilates from the GPP drop) and reduced heterotrophic respiration (high soil-water deficit over-compensating the built-in effect of warmer temperatures increasing decomposition according to Q_{10} relationships). (Q_{10} is the exponential factor of the temperature dependency of respiration.) Complementary measurements at Hesse confirm that pronounced soil water deficits compensated for the effect of warmer temperatures in reducing soil respiration. The parallel reduction in GPP and TER over western Europe ($4.6 \times 10^6 \text{ km}^2$) equates to a 2003 anomalous source to the atmosphere of 0.5 Pg C yr^{-1} , roughly undoing four years of net carbon storage⁶, although uncertainties in each number remain quite large (see Methods).

Finally, we analysed the 1900–2003 simulation to place the 2003 productivity reduction in context (Fig. 3). We find that 2003 has the lowest productivity of the past century (20% below the 1960–1990 average). This result is in good general agreement with crop harvest historical data¹³, especially for the summer crop maize. Such a crash in productivity during one year is large enough to alter the ‘greening’ detected from satellites, and interpreted as a mean NPP increase of 1% per year over Europe¹⁸.

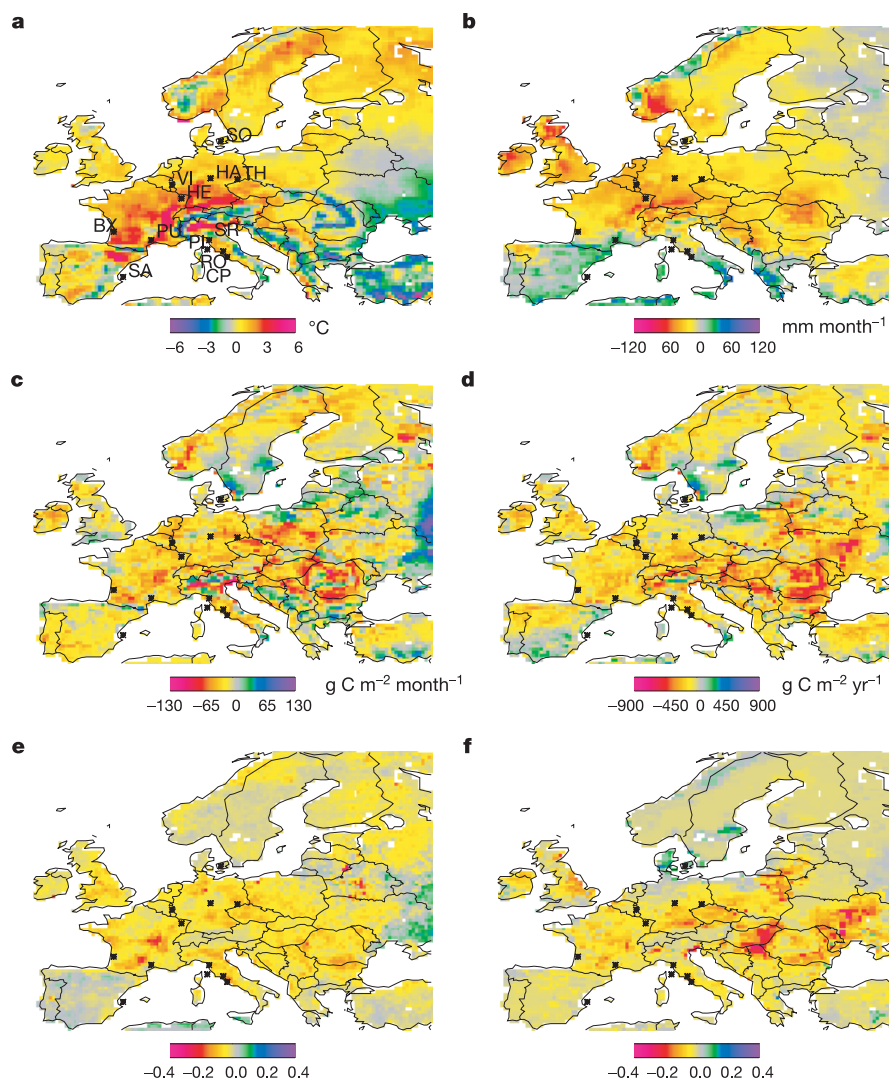


Figure 2 | European-wide anomalies of climate and net primary productivity (NPP) during 2003. All data compare 2003 and the average of 1998–2002. **a, b**, Climate. **a**, Changes in July–September air temperature. **b**, Changes in annual precipitation. **c, d**, NPP. **c**, Simulated changes in July–September NPP. **d**, Simulated changes in annual mean NPP.

e, f, Fraction of absorbed photosynthetic radiation. **e**, Observed changes in FAPAR from the MODIS-Terra-EOS satellite. **f**, Simulated changes in FAPAR. The location of eddy covariance sites is indicated by the black squares (see list in Table 1).

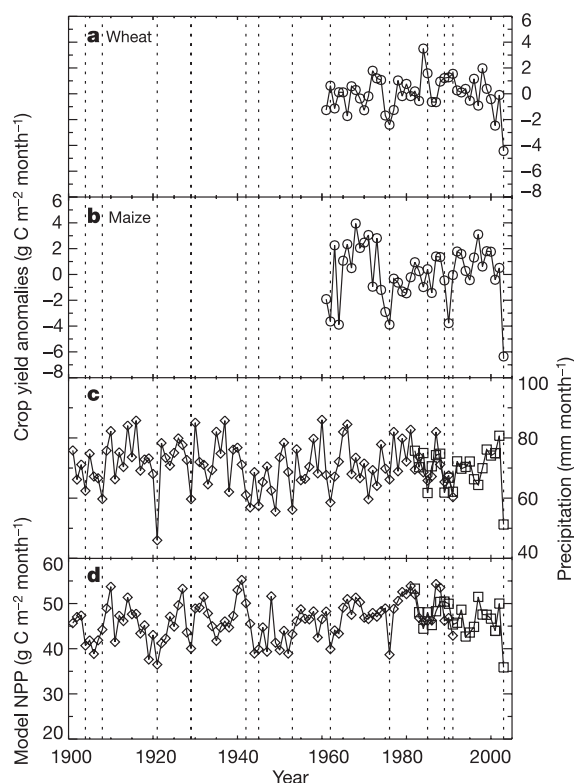


Figure 3 | Observed crop yield and modelled crop NPP changes in response to climate variability over France and Italy during the past 100 years.

a, Winter wheat yields. The trace shows area-weighted national yield records from ref. 13; a linear trend has been removed from the data to subtract the effects of improved agriculture and let appear the climate-induced variability. **b**, Same for maize. **c**, Annual precipitation over the same domain: diamonds are precipitation data from ref. 15 and squares from ref. 18. **d**, Model-simulated NPP obtained by averaging all cropland grid points in France and Italy. Dashed vertical lines indicate the driest years of the past 100 years.

In conclusion, extreme events such as the 2003 European drought and heatwave have the potential to significantly alter long-term continental carbon balances. In Europe, more frequent extreme drought events^{3,4} may counteract the effects of the anticipated mean warming and lengthening of the growing season, and erode the health and productivity of ecosystems, reversing sinks to sources, and contributing to positive carbon-climate feedbacks. This study only attempts to quantify the short-term consequences of extreme climate conditions on productivity (Supplementary Figs S1 and S5), but the long-term impacts are likely to be significant as well^{19,20}. In particular, we need to understand better the consequences of xylem embolism²¹, the effects of reduced carbohydrate pool sizes²² on subsequent leaf and fine root production and turnover, and on the ability of plants to resist pathogen attacks²³, the impacts on soil microbial dynamics, decomposition and nutrient-supply processes²⁴, and shifting competitive abilities between plant species.

METHODS

Eddy covariance data. CO₂ fluxes were determined by the eddy-covariance technique using the same protocol as in ref. 8. We used quality-controlled half-hourly data screened out site-specifically for low nocturnal turbulence²⁵ (u^* -filtering). The choice of 2002 as a reference period is justified by the availability of high-quality data from numerous sites, and because 2002 was not a very abnormal year over most of Europe. For the July–September period, 2002 was not warmer than the average 1998–2002 at any site, but it was wetter in central Europe and northern Italy (Tharandt, Hainich, San Rossore) and drier at two southern sites (Bray, El Saler). The flux data were gap-filled⁹ by replacing missing values with average measurements obtained under similar

meteorological conditions and during as small a time window as possible. The partitioning of NEE into GPP and TER was achieved through an algorithm⁹ that first establishes a short-term temperature dependence of TER from turbulent night-time data and then uses this relationship for extrapolating respiration from night-time to day-time. Day-to-day varying base rates of respiration were derived from u^* -filtered night-time fluxes, avoiding the confounding effect of covariance between general biological activity and temperature (<http://gaia.agraria.unitus.it/database/eddyproc/>). Uncertainties in the changes of GPP and TER between the years were estimated as a combination of errors arising from u^* -filtering, gap-filling and flux-partitioning. We assume that potential systematic errors affecting the absolute magnitude of the fluxes, as well as biases due to u^* -filtering cancel out by differencing between the years, because fluxes in both years should be affected similarly. Random errors of up to 50% for half-hourly fluxes diminish by integration over a month or a year. The bias of the gap-filling was estimated by introducing artificial gaps and was never higher than $3 \text{ g C m}^{-2} \text{ month}^{-1}$. At most 20% of the data were gap-filled, so this corresponds to an error of $0.6 \text{ g C m}^{-2} \text{ month}^{-1}$. The uncertainty of the GPP versus TER partitioning is largely determined by the uncertainty of the temperature sensitivity (E_0) used to extrapolate from night to day. This uncertainty was estimated as the standard deviation of all E_0 estimates for one year in ref. 9, assuming that the true value of E_0 is constant over the year and all variability can be attributed to the estimation error. Clearly, because E_0 can vary through the year, this is a conservative estimate of error. Errors for each year per site were summed for the difference between years, assuming that they are independent (Table 1).

Carbon flux model. Crop harvest-yield data are reported on a yearly basis¹³. Crop-specific factors¹¹ were used to convert harvested biomass into dry matter ($\sim 0.8 \text{ g}$), into carbon mass ($\sim 0.45 \text{ gC}$) and into crop NPP. The ORCHIDEE biosphere model⁵ (<http://www.ipsl.jussieu.fr/~ssipl/>) explicitly calculates CO₂, energy and H₂O fluxes on a half-hourly basis, forced by air temperature, precipitation, air humidity and solar and thermal radiation data. Turbulent fluxes are coupled to the calculation of carbon pool dynamics internal to the ecosystem, for eight different plant functional types over Europe. Carbon dynamics include calculation of the growth onset and senescence periods, allocation of assimilates to the different plant organs, mortality, and litter and soil organic matter decomposition. No explicit nitrogen cycle is modelled. No adjustments are made to adapt the modelled response of the vegetation to extreme climate conditions. For point-wise simulations at eddy-covariance sites, the modelled carbon fluxes and stocks are spun up (200 years) to their equilibrium value, and then forced by *in-situ* meteorological records. By construction, the long-term modelled NEE is zero, unlike in the flux measurements. Yet, interannual variations of NEE, GPP and TER in response to climate can be safely analysed. For spatially explicit simulations of carbon and water fluxes, a vegetation map^{26,27}, translated into the eight plant functional types, was held constant and ambient atmospheric CO₂ was increased from 280 to 360 p.p.m. between 1900 and 2003. After 1,000 yr spin-up, ORCHIDEE was driven from 1900 to 2003 by monthly mean historical climate reconstructions over Europe¹⁵ interpolated to half-hourly values using a weather generator over the period 1900–1990 at a spatial resolution of $\sim 100 \text{ km}$, six-hourly meteorological re-analyses¹⁶ provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) at $\sim 125 \text{ km}$ spatial resolution over the period 1983–2001, and six-hourly ECMWF operational weather analyses at $\sim 40 \text{ km}$ spatial resolution over the period 2000–2003. Using distinct climate input data sets is necessary to span over the whole period of interest, but creates some differences in the simulated carbon quantities. We verified that the year-to-year variability of CO₂ fluxes was, however, similar ($\pm 10\%$) between simulations forced by distinct climate data sets during their period of overlap. Using another climate data set (NCEP), we found a difference of up to 50% in the modelled NPP reduction and of 5% in the NEE reduction. To evaluate the model results against satellite FAPAR, we converted leaf area index into FAPAR following ref. 28.

Received 30 March; accepted 23 June 2005.

- Cox, P. M., Betts, R. A., Jones, C. D., Spal, A. S. & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).
- Dufresne, J. L. et al. On the magnitude of positive feedback between future climate change and the carbon cycle. *Geophys. Res. Lett.* **29**(10), doi:10.1029/2001GL013777 (2002).
- Schär, C. et al. The role of increasing temperature variability in European summer heatwaves. *Nature* **427**, 332–335 (2004).
- Meehl, G. A. & Tebaldi, C. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science* **305**, 994–997 (2004).
- Krinner, G. et al. A dynamic global vegetation model for studies of the

- coupled atmosphere-biosphere system. *Glob. Biogeochem. Cycles* **19**, 1–33 (2005).
6. Janssens, I. A. *et al.* Europe's terrestrial biosphere absorbs 7 to 12% of European anthropogenic CO₂ emissions. *Science* **300**, 1538–1542 (2003).
 7. Valentini, R. *et al.* Respiration as the main determinant of carbon balance in European forests. *Nature* **404**, 861–865 (2000).
 8. Aubinet, M. *et al.* Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology. *Adv. Ecol. Res.* **30**, 113–175 (2000).
 9. Reichstein, M. *et al.* On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Glob. Change Biol.* **11**, 1–16 (2005).
 10. Granier, A., Bréda, N., Piron, P. & Vilette, S. A lumped water balance model to evaluate duration and intensity of drought constraints in forest stands. *Ecol. Modell.* **116**, 269–283 (1999).
 11. Reichstein, M. *et al.* Inverse modelling of seasonal drought effects on canopy CO₂/H₂O exchange in three Mediterranean Ecosystems. *J. Geophys. Res.* **108**, 4716–4721 (2003).
 12. Irvine, J., Perks, M. P., Magnani, F. & Grace, J. The response of *Pinus sylvestris* to drought: stomatal control of transpiration and hydraulic conductance. *Tree Physiol.* **18**, 393–402 (1998).
 13. Food and Agriculture Organization Database (<http://faostat.fao.org/faostat/collections?subset=agriculture>) (2004).
 14. Goudriaan, J., Groot, J. J. R. & Uithol, P. W. J. in *Terrestrial Global Productivity* (eds Saugier, B. & Roy, J.) (Academic, 2001).
 15. Mitchell, T. D., Carter, T. R., Jones, P. D., Hulme, M. & New, M. A comprehensive set of high resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100). Working paper 55 (Tyndall Centre for Climate Change Research, July 2004); available at (http://www.tyndall.ac.uk/publications/working_papers/wp55.pdf).
 16. Simmons, A. J. & Gibson, J. K. *The ERA 40 Project Plan*. ERA-40 Project Report Series No. 1, 1–62 (European Center for Medium Range Weather Forecasts (ECMWF), 2000).
 17. Myneni, R. B. *et al.* Global products of vegetation leaf area and fraction absorbed PAR from one year of MODIS data. *Remote Sens. Environ.* **83**(1–2), 214–231 (2002).
 18. Nemani, R. R. *et al.* Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* **300**, 1560–1563 (2003).
 19. Barber, V. A. *et al.* Reduced growth of Alaskan white spruce in the twentieth century from temperature-induced drought stress. *Nature* **405**, 668–673 (2000).
 20. Elliot, K. J. & Swank, W. T. Impacts of drought on tree mortality and growth in a mixed hardwood forest. *J. Veget. Sci.* **5**, 229–236 (1994).
 21. Tyree, M. T. & Sperry, J. S. Do woody plants operate near the point of catastrophic xylem dysfunction caused by dynamic water stress? Answers from a model. *Plant Physiol.* **88**, 574–580 (1988).
 22. Cherbuy, B., Joffre, R., Gillon, D. & Rambal, S. Internal remobilization of carbohydrates, lipids, nitrogen and phosphorus in the Mediterranean evergreen oak *Quercus ilex*. *Tree Physiol.* **21**, 9–17 (2001).
 23. Boyer, J. S. Biochemical and biophysical aspects of water deficits and the predisposition to disease. *Annu. Rev. Phytopathol.* **33**, 251–274 (1995).
 24. Schimel, J. S. *et al.* Moisture effects on microbial activity and community structure in decomposing birch litter in the Alaskan taiga. *Soil Biol. Biochem.* **31**, 831–838 (1999).
 25. Reichstein, M. *et al.* Severe drought effects on ecosystem CO₂ and H₂O fluxes at three mediterranean sites: revision of current hypothesis? *Glob. Change Biol.* **8**, 999–1017 (2002).
 26. Büttner, G., Feranec, J. & Jaffrain, G. *Corine Land Cover Update 2000*. Technical Report 89 (European Environment Agency, 2000); available at (http://reports.eea.eu.int/technical_report_2002_89/en).
 27. Mucher, C. A., Steinnocher, K. T., Kressler, F. D. & Heunks, D. Land cover characterization and change detection for environmental monitoring of pan europe. *Int. J. Remote Sens.* **21**, 1159–1182 (2000).
 28. Monsi, M. & Saeki, T. Über den Lichtfaktor in den Pflanzengesellschaften und seine Bedeutung für die Stoffproduktion. *Jpn. J. Bot.* **14**, 22–52 (1953).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is part of the CARBOEUROPE-IP research program funded by the European Union. Eddy-covariance flux towers are also funded by national programmes. M.R. is supported by a European Union Marie-Curie Fellowship. Supercomputing facilities are provided by the French Commissariat à l'Energie Atomique. We thank H. Dolman, M. Heimann and J. Grace for discussions while preparing this manuscript.

Author Contributions The first three authors contributed equally to this work: Ph.C. did the analysis, M.R. did the eddy covariance data harmonization and interpretation, and N.V. did the modelling.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Ph.C. (philippe.ciais@cea.fr).

LETTERS

Minimum speed limit for ocean ridge magmatism from ^{210}Pb – ^{226}Ra – ^{230}Th disequilibria

K. H. Rubin¹, I. van der Zander¹, M. C. Smith^{1†} & E. C. Bergmanis¹

Although 70 per cent of global crustal magmatism occurs at mid-ocean ridges¹—where the heat budget controls crustal structure, hydrothermal activity and a vibrant biosphere—the tempo of magmatic inputs in these regions remains poorly understood. Such timescales can be assessed, however, with natural radioactive-decay-chain nuclides, because chemical disruption to secular equilibrium systems initiates parent–daughter disequilibria, which re-equilibrate by the shorter half-life in a pair. Here we use ^{210}Pb – ^{226}Ra – ^{230}Th radioactive disequilibria and other geochemical attributes in oceanic basalts less than 20 years old to infer that melts of the Earth's mantle can be transported, accumulated and erupted in a few decades. This implies that magmatic conditions can fluctuate rapidly at ridge volcanoes. ^{210}Pb deficits of up to 15 per cent relative to ^{226}Ra occur in normal mid-ocean ridge basalts, with the largest deficits in the most magnesium-rich lavas. The 22-year half-life of ^{210}Pb requires very recent fractionation of these two uranium-series nuclides. Relationships between ^{210}Pb deficits, ($^{226}\text{Ra}/^{230}\text{Th}$) activity ratios and compatible trace-element ratios preclude crustal-magma differentiation or daughter-isotope degassing as the main causes for the signal. A mantle-melting model² can simulate observed disequilibria but preservation requires a subsequent mechanism to transport melt rapidly. The likelihood of magmatic disequilibria occurring before melt enters shallow crustal magma bodies also limits differentiation and heat replenishment timescales to decades at the localities studied.

^{210}Pb – ^{226}Ra disequilibria can set a lower temporal bound on magmatic chemical fractionation because the ^{210}Pb half-life is among the shortest of petrologically applicable actinide series nuclides^{3–9}. The focus of prior studies (all but one involving subaerial volcanoes) has been on crustal magmatic timescales. Observed 10–20% ^{210}Pb deficits have been hypothesized to form by shallow magmatic degassing of ^{222}Rn (an intermediate nuclide with a half-life of 3.8 days that is produced on the ^{226}Ra decay path to ^{210}Pb), either by degassed/undegassed magma mixing⁴ or steady-state ^{222}Rn loss from convecting magma⁸ over less than 100 years, or by recent ^{226}Ra contamination⁶. However, alkalic Vestmannaeyjar (Iceland) lavas have 15–33% ^{210}Pb deficits that diminish with increased differentiation and are consistent with closed-system ingrowth⁷, implying a non-crustal source. Pb degassing is unlikely to induce significant ^{210}Pb – ^{226}Ra disequilibria because it is only slightly volatile at magmatic conditions (1–2% magmatic abundance loss upon eruption¹⁰). ^{222}Rn accumulation^{3,9}, young volcanic sublimate contamination⁸, plagioclase accumulation⁵ or rainwater alteration⁵ are some hypothetical sources of less common ^{210}Pb -excesses.

Sampling difficulty and largely unknown deep-sea eruption histories have inhibited equivalent mid-ocean ridge basalts (MORB) ^{210}Pb – ^{226}Ra studies. This first systematic investigation uses samples that were collected from a manned submersible, from well-charac-

terized, known-age, recent eruptions. Sample localities cover a broad range of magmatic conditions (see lava flow, sample and analytical details in Methods). ^{210}Pb – ^{226}Ra – ^{230}Th varies systematically within the sample suite (Figs 1 and 2) but the following discussion focuses on normal (n)-MORB data, concluding with contrasting behaviour at the Loihi and Axial seamounts. n-MORB ^{210}Pb – ^{226}Ra and ^{230}Th – ^{238}U disequilibria are of similar magnitude (0–15% and 10–20%, respectively); as in other MORB^{11,12}, ^{226}Ra – ^{230}Th disequilibria are much larger (100–200%), with greatest values in the lowest-U-abundance samples¹³. Our MORB data overlap a previously noted inverse ($^{230}\text{Th}/^{238}\text{U}$) versus ($^{226}\text{Ra}/^{230}\text{Th}$) relationship, best expressed in 9°–10° N East Pacific Rise (EPR) lavas (ref. 12 and therein), where MgO covariations were also noted (see below). Good inverse ($^{210}\text{Pb}/^{226}\text{Ra}$) correlations with ($^{226}\text{Ra}/^{230}\text{Th}$), Mg content, Ni/Co (and other compatible element ratios that decrease during olivine $\pm$ clinopyroxene fractionation), and Sr/Y (which decreases during shallow plagioclase crystallization) occur with one side pinned at ($^{210}\text{Pb}/^{226}\text{Ra}$) = 1. There is no simple ($^{210}\text{Pb}/^{226}\text{Ra}$) relationship to ($^{230}\text{Th}/^{238}\text{U}$) or incompatible element ratios sensitive to source composition (for example, Th/U). A strong Ni/V–Sr/Y relationship (Fig. 2h) suggests primarily a crustal differentiation signature. Intra-lava-flow compatible element covariations mimic overall n-MORB trends: larger coupled ^{210}Pb – ^{226}Ra – ^{230}Th disequilibria in most mafic and lowest-U-abundance samples.

How ^{210}Pb – ^{226}Ra disequilibria constrain magmatic timescales at submarine ridge and seamount volcanoes depends on which of numerous potential sources operate. The inter-correlation of n-MORB ^{210}Pb – ^{226}Ra – ^{230}Th –Mg compositions implies a common cause. Different magnitude ($^{210}\text{Pb}/^{226}\text{Ra}$) and ($^{226}\text{Ra}/^{230}\text{Th}$) variations rule out primary ^{226}Ra addition to an equilibrium system. The highest n-MORB ^{210}Pb -deficits occur at lowest (^{226}Ra), precluding secondary ^{226}Ra addition to a system initially at ($^{226}\text{Ra}/^{230}\text{Th}$) > 1 and ($^{210}\text{Pb}/^{226}\text{Ra}$) = 1. Rather, the positive (^{210}Pb) versus ($^{210}\text{Pb}/^{226}\text{Ra}$) correlation suggests relative ^{210}Pb loss controls ^{210}Pb – ^{226}Ra disequilibria (Fig. 2).

As at Vestmannaeyjar⁷, we can exclude a shallow differentiation origin of n-MORB ^{210}Pb deficits because the largest occur in the least-fractionated magmas and those with the largest ^{226}Ra excesses (a MORB signature generally accepted to result primarily from mantle melting^{11,12}), and because no positive ($^{210}\text{Pb}/^{226}\text{Ra}$) versus Sr/Y trend is observed. Likewise, ^{222}Rn degassing is an unlikely mechanism because effects should be maximized at longer residence times in limited recharge convecting magma chambers⁸ (that is, in the most differentiated co-magmatic samples). Even if ^{222}Rn degassing were more rapid than differentiation, we would not anticipate the greatest ^{210}Pb deficit in magmas with the shortest implied crustal residence times (the most magnesium-rich).

Immiscible Fe–Ni sulphides (which can incorporate Pb) are

¹Department of Geology and Geophysics, Hawaii Center for Volcanology, University of Hawaii, 1680 East-West Road, Honolulu, Hawaii 96822, USA. [†]Present address: Department of Geological Sciences, University of Florida, Gainesville, Florida 32611, USA.

another potential source of magmatic ^{210}Pb deficits; however, ^{210}Pb – ^{226}Ra covariations with compatible elements (Fig. 2) do not follow their sulphide compatibility sequence ($\text{Ni} > \text{Cu} > \text{Co} > \text{V}$), from which we would predict the greatest deficits at the lowest Ni/X ratios.

Diffusion between lower crustal gabbroic cumulates at chemical disequilibrium with initially secular equilibrium melts is hypothetically capable of inducing U-series disequilibria under some melt-percolating conditions¹³ (for example, sufficiently fast solid–melt interaction, juxtaposition of chemical disequilibrium melts and residues, static grain boundaries, low porosity). Solid-state diffusivity and the mineral–melt distribution coefficient (D_i) differences in cumulate clinopyroxene and plagioclase can drive Ra–Th fractionation to produce melt ^{226}Ra – ^{230}Th disequilibria¹³ with little effect on ($^{230}\text{Th}/^{238}\text{U}$). That study¹³ (which did not consider ^{210}Pb) found that modelled clinopyroxene-rich cumulates create ^{226}Ra -excesses more easily than plagioclase cumulates (for many situations the latter induce melt ^{226}Ra -deficits). Pb/Ra diffusivity is 1 in clinopyroxene¹³ so this mineral cannot diffusively fractionate Pb from Ra. Pb/Ra diffusivity is $\sim 10^4$ in plagioclase^{13,14}, so diffusion of ^{226}Ra to, or ^{210}Pb from, a melt could induce a melt deficit. However, because both elements have $\text{plag}D_i < 1$ (ref. 15), diffusive conditions favouring melt ^{226}Ra -excesses also favour ^{210}Pb migration to the melt (at a much faster rate); thus, Ra and Pb must flow in the same direction between initially secular equilibrium melts and solids. This, combined with the above-noted inability of Ra addition to produce the n-MORB ^{210}Pb – ^{226}Ra – ^{230}Th disequilibria and of clinopyroxene to diffusively fractionate Pb and Ra, imply that diffusive fractionation did not primarily cause our disequilibria. But if it did, disequilibria preservation would limit subsequent upper crustal magma residence time to $< 4\text{--}5 \times$ the ^{210}Pb half-life once melts segregate from the cumulate mush.

It can be argued that some or all of the observed disequilibria are produced in the mantle. Mantle melting/transport can induce a negative ($^{210}\text{Pb}/^{226}\text{Ra}$)–($^{226}\text{Ra}/^{230}\text{Th}$) correlation (Fig. 3) because D_i values¹⁵ favour such partitioning (see Methods). The difficulty comes in preserving this signature once melts segregate from the

melting regime, because it will have largely decayed after a combined 110-yr ($5 \times$ the ^{210}Pb half-life) of magma transport and crustal residence (Fig. 3). Single porosity instantaneous melts² do not easily match the steep n-MORB ($^{210}\text{Pb}/^{226}\text{Ra}$) versus ($^{226}\text{Ra}/^{230}\text{Th}$) trend except at unrealistically low D_{Pb} , but mixed melts after different amounts of closed-system decay do (Fig. 3).

McKenzie¹⁶ argues that near-fractional melts are segregated and transported to the lithosphere from all mantle melt column depths on timescales of less than the ^{226}Ra half-life (for example, tens of kilometres in hundreds to thousands of years). ^{210}Pb – ^{226}Ra disequilibria would require $\sim 75 \times$ faster transport times ($\leq 10 \text{ km yr}^{-1}$), most probably necessitating a system of self-sustaining channels and dikes. However, such melting does not explain the inverse n-MORB ($^{210}\text{Pb}/^{226}\text{Ra}$) trends versus ($^{226}\text{Ra}/^{230}\text{Th}$), Mg# and Ni/X ratios. ^{210}Pb decay during shallow differentiation from a larger, uniform, initial ^{210}Pb deficit can explain aspects of the variations but the timescale is too short to modify ($^{226}\text{Ra}/^{230}\text{Th}$) as well. Mixing of older, shallowly differentiated low ^{226}Ra -excess and ($^{210}\text{Pb}/^{226}\text{Ra}$) = 1 melts with younger, more mafic melts (that is, at larger ^{210}Pb – ^{226}Ra – ^{230}Th disequilibria) can explain the trends. Then, the largest and smallest ^{210}Pb -deficit magmas would reflect, respectively, fast melting and transport timescales (for example, 44–66 yr, Fig. 3) and superimposed mixing and crustal ageing ($\sim 200\text{--}400$ yr, from ($^{226}\text{Ra}/^{230}\text{Th}$) differences).

Another possibility produces ^{210}Pb – ^{226}Ra – ^{230}Th variations primarily within the mantle. ($^{230}\text{Th}/^{238}\text{U}$) and ($^{226}\text{Ra}/^{230}\text{Th}$) versus MgO trends in northern-EPR MORB¹² have been proposed to reflect melt mixing from different mantle depths and two transport mechanisms^{12,17}. Low-porosity chromatographic flow delivers high- ^{226}Ra -excess, low- ^{230}Th -excess melts to the shallow mantle, and high-porosity unreactive channels transport high ^{230}Th -excess, ($^{226}\text{Ra}/^{230}\text{Th}$) = 1 melts formed by porous flow deeper in the mantle¹⁷. Different major-element compositions (for example, MgO variations) ensue from polybaric melting and subsequent crystallization. The more magnesium-rich shallow porous flow fraction could contain ^{210}Pb deficits for much the same reason as do mckenziean fractional melts. Almost all magmas reaching the crust

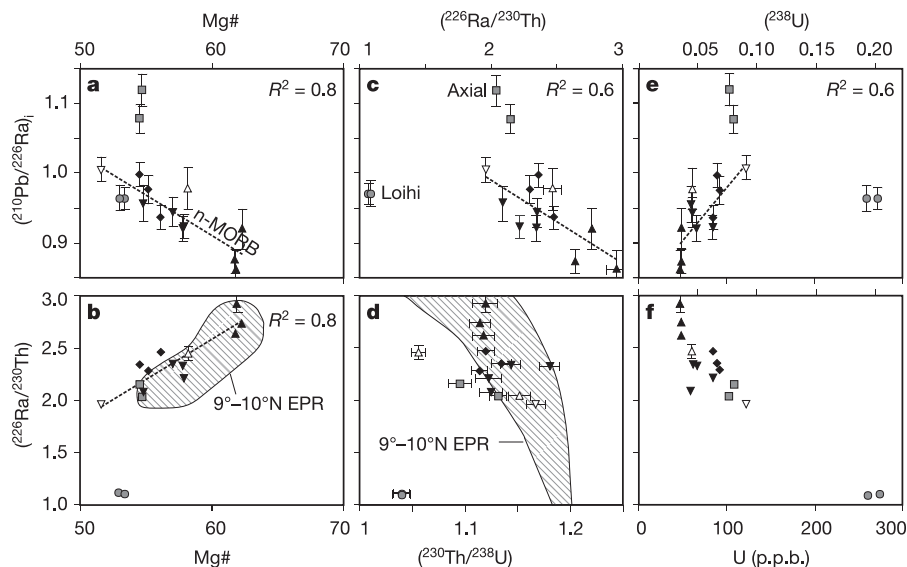


Figure 1 | U-series data for young lava flows. Symbols are as follows. Diamonds, JdF; downtriangles, southern EPR; uptriangles, northern EPR; circles, Loihi; squares, Axial. Black, grey and open symbols are n-MORB, seamount, and older n-MORB lava flows, respectively; see Table 1. **a–b**, ^{210}Pb – ^{226}Ra – ^{230}Th activity ratios versus Mg#; **c–d**, Activity ratio interrelationships; **e–f**, ^{210}Pb – ^{226}Ra – ^{230}Th activity ratios versus U abundance. n-MORB data are strongly correlated (linear regressions exclude

seamount data) in all plots except in **d**, where southern-EPR data contravene an otherwise vertical trend. The largest ^{210}Pb – ^{226}Ra – ^{230}Th disequilibria occur in the most-magnesium-rich and lowest-U-abundance lavas. Error bars reflect 2σ s.e. statistics for all nuclides except ^{210}Pb (1 σ s.e.) and are smaller than the data symbols when not shown (see Methods for error estimation details). Stippled fields are taken from ref. 12.

are mixtures in this model, but ^{210}Pb – ^{226}Ra disequilibria would still require a significant segregated melt fraction to experience upper-mantle and crustal transport in a few decades. This model partitions more time and compositional variation to the mantle, and less to the crust, compared to the fractional melting/differentiation/crustal-mixing scenario. Although either or both situations might operate, a plagioclase role in the Sr/Y versus Ni/V versus ($^{210}\text{Pb}/^{226}\text{Ra}$) trend favours the latter.

MORB ^{210}Pb – ^{226}Ra disequilibria also illuminate crustal magma dynamics. If magmas enter the crust with ^{210}Pb -deficits and decay to ($^{210}\text{Pb}/^{226}\text{Ra}$) = 1 during shallow differentiation, then the very presence of disequilibria implies short residence times in eruptible melt lenses (at most two to three decades). It has been hypothesized^{18–20} that weaker/thinner crust at high magma supply ridges (common at high spreading rates) promotes small, frequent eruptions, whereas longer magma accumulation times produce larger and less frequent eruptions through stronger/thicker crust (for example, at lower spreading rates). ^{210}Pb -deficit patterns are consistent with

this: slightly smaller deficits in Juan de Fuca relative to EPR MORB suggest longer melt accumulation rates, and slightly less variable southern-EPR deficits (3.5% range, versus 7% at the northern EPR and Juan de Fuca) imply the most uniform rates. Additionally, the large Aldo-Kihi Th/U range (7%) at limited ($^{210}\text{Pb}/^{226}\text{Ra}$) variation indicates that a mixed-source magma experienced faster and less homogenizing melt transport/differentiation conditions, whereas the significant ($^{210}\text{Pb}/^{226}\text{Ra}$) variation at nearly constant Th/U in the North Cleft Sheet (2.1%) and BBQ (0.5%) flows reflect compositionally more homogeneous sources and more variable melt accumulation rates.

The lack of large seamount ^{210}Pb deficits provides additional constraints. Mantle melting cannot easily generate ^{210}Pb excesses (for example, at Axial seamount) using published D_{Pb} and D_{Ra} ¹⁵ (Fig. 3). Instead, excesses probably reflect processes (including but not limited to ^{222}Rn or plagioclase⁵ accumulation) occurring in its large, long-lived crustal magma chamber (for example, $\leq 2\%$ of stored magma²¹ erupted in 1998). At equilibrium, (Pb/Ra)_{plag} and (Sr/Y)_{plag} can be 10–100 × that in the coexisting melt¹⁵. Although plagioclase removal did not produce Aldo-Kihi ^{210}Pb – ^{226}Ra disequilibria (^{210}Pb deficits decrease to equilibrium as Sr/Y decreases),

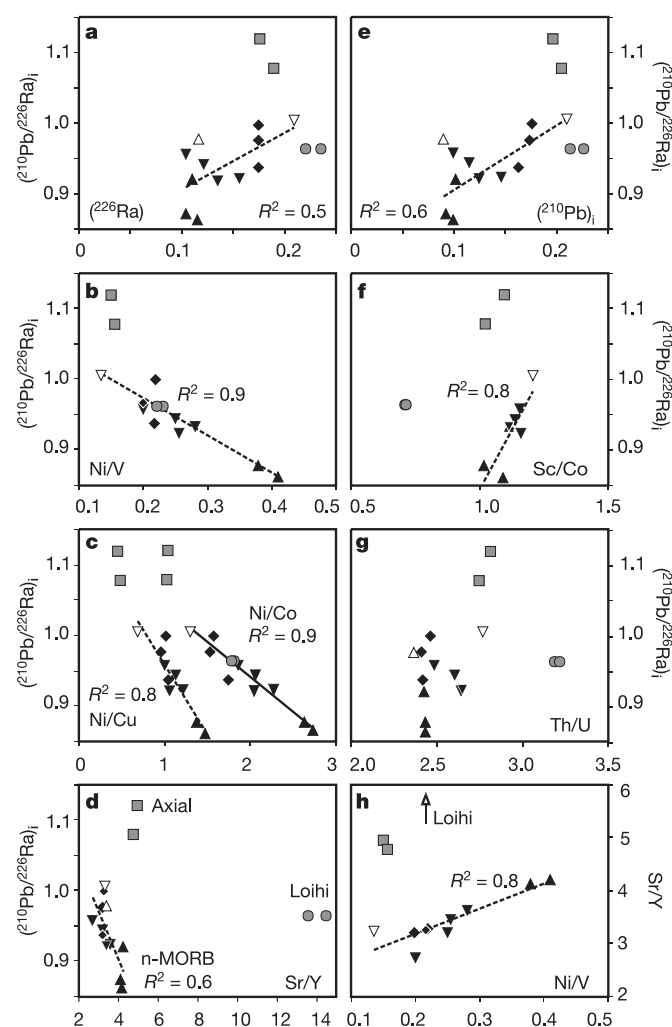


Figure 2 | ^{210}Pb – ^{226}Ra radioactive disequilibria versus other chemical parameters. Regressions include n-MORB data. **a, e**, ($^{210}\text{Pb}/^{226}\text{Ra}$) versus (^{210}Pb) and (^{226}Ra) consistent with relative Pb subtraction but not relative Ra addition for n-MORB disequilibria. **b–d, f**, ($^{210}\text{Pb}/^{226}\text{Ra}$) correlations indicate that the least olivine + plagioclase $\pm$ clinopyroxene fractionated lavas have the greatest disequilibria. **g**, No ($^{210}\text{Pb}/^{226}\text{Ra}$) overall correlation with Th/U (a source enrichment indicator); we note the vertical North Cleft Sheet and northern-EPR arrays in panels **d** and **g** and greater dispersion of southern-EPR Aldo-Kihi lavas (see text). **h**, Correlation shows plagioclase control on fractionation. Some elemental data is unavailable in some panels; symbols as in Fig. 1. ($^{210}\text{Pb}/^{226}\text{Ra}$) errors are omitted for clarity.

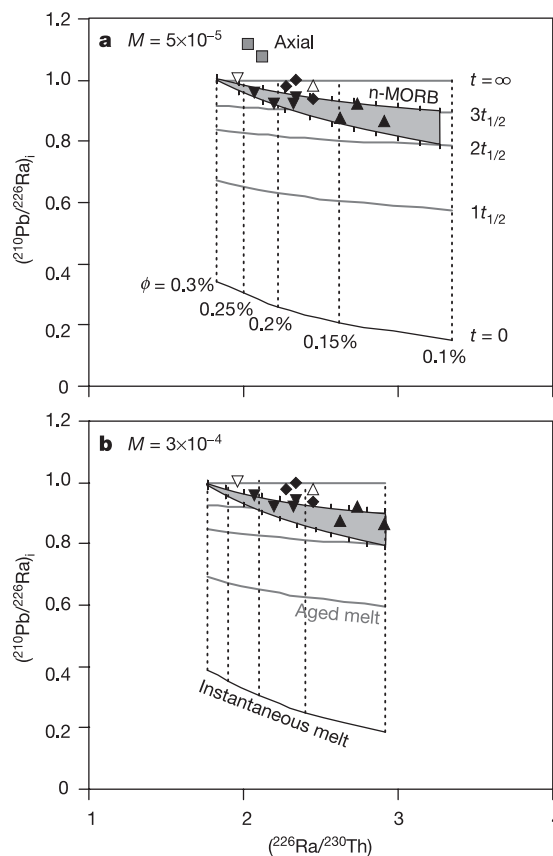


Figure 3 | Instantaneous melting² and ^{210}Pb – ^{226}Ra – ^{230}Th disequilibria. **a, b**, Melting rates (M) bracket sample ^{226}Ra – ^{230}Th – ^{238}U (low M produces greater ^{230}Th – ^{238}U disequilibrium). Substantial ^{210}Pb deficits ensue for a broad range of melting conditions (just a few simulations are shown). Vertical dashed lines show calculations at different porosities, ϕ , in 0.5% increments (smaller ϕ values produce greater ^{226}Ra – ^{230}Th disequilibria). Melt compositions after closed-system decay are shown in lighter grey. We note that the melting trajectory shallows during decay towards MORB ($^{210}\text{Pb}/^{226}\text{Ra}$). The steeper ($^{210}\text{Pb}/^{226}\text{Ra}$) versus ($^{226}\text{Ra}/^{230}\text{Th}$) data trend is mimicked by instantaneous melting with $D_{\text{Pb}} = 2\text{--}3 \times D_{\text{Ra}}$, but such D_{Pb} are $\leq 100 \times$ the literature values¹⁵. Grey swathes represent magma-chamber mixing of primitive small degree melts 44–66 yr after segregation with 220-yr-old melts (other temporal combinations are possible). Symbols as in Fig. 1. $t_{1/2}$ is ^{210}Pb half life.

Table 1 | Geochemical data for young submarine lava flows

Location	Eruption	Sample	Eruption or correction date	²¹⁰ Pb analysis date	(²³⁰ Th)/(²³⁸ U)	(²²⁶ Ra)/(²³⁰ Th)	(²¹⁰ Pb)/(²²⁶ Ra)	(²¹⁰ Pb)/(²²⁶ Ra) _i	Mg#
Loihi	1996 eruption ²⁴	P286-6b	30 Jun 96*	4 Sep 02	1.039 ± 0.010	1.118 ± 0.012	0.970 ± 0.017	0.964 ± 0.019	53.0
		P287-1	30 Jun 96*	16 Aug 02	1.039 ± 0.010	1.103 ± 0.011	0.970 ± 0.014	0.964 ± 0.016	53.4
Juan de Fuca North Cleft	Young sheet flow ^{18,19,29}	2077-3a	1 Jan 85†	16 Aug 02	1.120 ± 0.010	2.465 ± 0.026	0.963 ± 0.015	0.937 ± 0.017	56.1
		2079-2	1 Jan 85†	27 Oct 00	1.134 ± 0.012	2.343 ± 0.026	0.999 ± 0.016	0.998 ± 0.018	54.5
		2092-1	1 Jan 85†	22 Feb 01	1.114 ± 0.009	2.288 ± 0.026	0.986 ± 0.019	0.976 ± 0.020	55.2
		R501-9	10 Jan 98‡	06 Sep 02	1.096 ± 0.013	2.150 ± 0.027	1.068 ± 0.018	1.078 ± 0.021	54.5
Juan de Fuca Axial seamount	1998 summit eruption ^{26,30}	R501-10	10 Jan 98‡	06 Jan 03	1.132 ± 0.009	2.040 ± 0.030	1.103 ± 0.021	1.120 ± 0.024	54.7
Northern EPR 9° 50' N	1991-2 Tubeworm BBQ flow ^{18,19,28}	2392-1	30 Dec 90*	04 Aug 00	1.117 ± 0.013	2.641 ± 0.039	0.906 ± 0.016	0.873 ± 0.018	61.8
		2504-1	2 Nov 91*	10 Aug 00	1.114 ± 0.012	2.742 ± 0.032	0.939 ± 0.030	0.920 ± 0.031	62.3
		2372-1	25 Feb 91*	24 Jul 00	1.120 ± 0.014	2.928 ± 0.086	0.897 ± 0.029	0.862 ± 0.030	61.9
		2502-1	2 Nov 91§	27 Oct 00	1.060 ± 0.009	2.457 ± 0.069	0.983 ± 0.029	0.977 ± 0.032	58.2
Southern EPR 17° 28' S	'Next oldest' unit ¹⁹	ND3-5	1 Dec 93	18 Aug 97	1.143 ± 0.011	2.346 ± 0.031	0.950 ± 0.020	0.944 ± 0.022	57.0
		ND19-8	1 Dec 93	08 Dec 97	1.181 ± 0.012	2.335 ± 0.025	0.930 ± 0.018	0.920 ± 0.020	57.8
		ND20-5	1 Dec 93	23 Nov 99	1.126 ± 0.014	2.086 ± 0.026	0.964 ± 0.025	0.957 ± 0.027	54.8
	'Distinct' unit ²⁰	ND6-1	1 Dec 93¶	23 Nov 99	1.123 ± 0.016	2.211 ± 0.033	0.936 ± 0.016	0.923 ± 0.018	57.9
		ND18-1	1 Dec 93§	26 Aug 97	1.168 ± 0.012	1.968 ± 0.024	1.005 ± 0.015	1.005 ± 0.017	51.6

All U-series nuclides were determined by high abundance sensitivity thermal ionization mass spectrometry, except ²¹⁰Pb (by α-counting). Parentheses denote activity ratios, which were calculated from atomic abundance analyses using $\lambda^{238}\text{U} = 1.551 \times 10^{-10} \text{ yr}^{-1}$, $\lambda^{230}\text{Th} = 9.195 \times 10^{-6} \text{ yr}^{-1}$ and $\lambda^{226}\text{Ra} = 4.332 \times 10^{-4} \text{ yr}^{-1}$ (²¹⁰Pb/²²⁶Ra) are given for both the analysis date and the eruption date (denoted by subscript 'i') and calculated using $\lambda^{210}\text{Pb} = 3.108 \times 10^{-2} \text{ yr}^{-1}$. Data acquisition and handling described in the Methods. Rocks were very fresh and mean (²³⁴U/²³⁸U) = 1.000 ± 0.002 (using $\lambda^{234}\text{U} = 2.835 \times 10^{-6} \text{ yr}^{-1}$); see Supplementary Table 1 for sample descriptions, compositional metadata and individual (²¹⁰Pb), (²²⁶Ra), (²³⁰Th), (²³⁴U/²³⁸U), Th, U and (²³⁰Th/²³²Th) data. See the references for geological and geochemical observations of these lava flows and for overall eruption details.

Mg# = (mol Mg)/(mol Mg) + (mol Fe²⁺) using $\Sigma\text{Fe} = \text{Fe}^{2+}$. The eruption or correction date is used to age-correct measured (²¹⁰Pb), as described in the notes below.

*Eruption date by ²¹⁰Po–²¹⁰Pb dating.

†Eruption date is the midpoint of the 4-yr interval between bathymetric surveys before and after the lava field emplacement.

‡Eruption date by seismic detection.

§Older lava unit based on field observations. Sample collection dates were used for age correction. For 2502-1 a 29–58-yr difference in time since ²¹⁰Pb–²²⁶Ra fractionation relative to the 1991–92 eruptive can be estimated from (²¹⁰Pb/²²⁶Ra)_i differences.

|| Sample collection by the submersible *Nautilus* during the 1993 *Naudur* expedition; geochemical, biological and geological evidence suggest an eruption occurred in 1991 or 1992, but physical variations within this lava flow also suggest a protracted eruption history²⁰ (additional details and references in Supplementary Table 1).

¶ Identified in the field as geologically and geographically distinct, and visually older than Aldo-Kihi. ²¹⁰Pb–²²⁶Ra disequilibrium is among the largest observed in Aldo-Kihi, suggesting very little, if any resolvable time difference since ²¹⁰Pb–²²⁶Ra fractionation for the two flows (less than a few years).

combined high Sr/Y at Axial (150% of the nearby North Cleft Sheet lava), minor ²¹⁰Pb deficits in the North Cleft Sheet flow, and sufficient residence time for melting signature decay all make accumulated plagioclase a potential source of small Axial ²¹⁰Pb excesses. Whether by plagioclase, Rn, or another magma chamber process, the lack of n-MORB excesses points to the evanescence of subridge melt reservoirs at these spreading/melt-supply rates. Rapid melt production in the mantle produces very limited ²²⁶Ra–²³⁰Th–²³⁸U disequilibria (10% of MORB values) in Hawaiian tholeiites^{22,23}, such that only small ²¹⁰Pb deficits would be expected from melting at Loihi. Seismicity, mineralogy and volatile contents suggest 1996 magma residence in a relatively deep (>9 km) reservoir beneath Loihi²⁴, perhaps hindering the formation of ²¹⁰Pb excesses.

Overall, detection of MORB ²¹⁰Pb–²²⁶Ra disequilibria is a milestone implying rapid mantle and crust magmatic timescales, despite uncertainty over the proportions of ²¹⁰Pb decay within each locale. Such disequilibria shave at least a factor of ten from magmatogenesis and melt transfer 'clocks', constrained in the 1980s to ~10⁵ years (ref. 25) and then ~10³ years (ref. 11) by ²³⁰Th–²³⁸U and ²²⁶Ra–²³⁰Th disequilibria, respectively.

METHODS

Five well-characterized and dated, Pacific Ocean submarine lava flow units were chosen because they were produced under a range of mantle and crustal conditions. The three n-MORB localities (Juan de Fuca, northern-EPR, southern-EPR) span a range of erupted volume, spreading rate (moderate to superfast) and compositional variation^{18–20} from passively upwelled mantle. Juan-de-Fuca-centred Axial seamount lavas reflect a MORB-like melting regime²⁶ with an atypical long-lived magma chamber²¹. Loihi seamount (Hawaii) lavas reflect buoyantly upwelling mantle^{22,24}. Multiple samples were studied from each site (Table 1 references provide geological and geochemical context).

Chemical analysis. Very fresh basaltic glasses were handpicked under a binocular microscope. Single sample dissolutions were split for various U-series analyses. High abundance sensitivity ion-counting thermal ionization mass spectrometry (TIMS) of all nuclides used a Sector54-warp^{23,27} except ²¹⁰Pb, which was by high-resolution ²¹⁰Po α-spectroscopy²⁸ (with a six-detector Canberra Alpha Analyst, efficiencies are 30.5% to 32.3%). See Supplementary

Table 1 for additional nuclide activity, activity ratio, and Th and U abundance data, and Supplementary Table 2 for major/trace element data.

U-series samples were processed in a clean laboratory using in-house ultraclean reagents. To minimize differences in analysis conditions, we optimized sample sizes (1–2 g) and split fractions to provide nearly constant analyte quantities: 50 ng, 25 fg, 0.5–1 ng and 0.025 c.p.m. for Th and U isotopic composition (IC), and ²²⁶Ra, Th–U and ²¹⁰Pb isotope dilution (ID) concentration measurements, respectively. Samples were completely dissolved using previously reported methods²³, split into Th–U ID, ²¹⁰Pb ID and Th–U IC plus ²²⁶Ra ID aliquots, and spiked with calibrated ²²⁹Th–²³³U, ²⁰⁹Po and ²²⁸Ra tracers, respectively. After ion exchange chromatography, analytes were loaded onto Re (Th, U) or W (Ra) ribbons for TIMS analysis^{23,27}. Po was auto-reduced onto Ag disks from pH = 2 HCl_{aq} with hydroxylamine hydrochloride and sodium citrate²⁸. Total procedural blanks were either below detection or <0.1% of all analyte abundances. Chemical yields were typically >90% for Th, 70–95% for ²²⁶Ra and ²¹⁰Pb, and 50–75% for U (additional details are in Table 1).

Analytical uncertainties. Reported uncertainties reflect fractional errors of data acquisition, weighing and spike calibration. TIMS analysis errors are 2σ standard errors of 100 to 400 ratios. Alpha-spectrometric analysis errors are 1σ standard errors of 2 or 3 ratios from separate acquisition periods over 1 to 2 months. Although slightly lower than more common estimates using counting statistics of total accumulated counts, the reproducibility of replicates was within stated error limits.

IC precision and accuracy were evaluated by repeat isotopic standard analysis throughout the study period: UCSC-A Th (*n* = 115), U010 (*n* = 12), and CRM112A U (*n* = 9), with respective relative precisions (2σ) of 1%, 0.4% and 0.1%. The respective IC accuracies were 0.3%, 0.01%, 0.06%, expressed as deviations between our means and accepted/certified values for those standards. A further reproducibility test of ²³²Th/²³⁰Th involved replicate analyses of 12 of 16 unknowns: 11 of 12 are within reported UCSC-A external precision (Supplementary Table 3).

ID precision and accuracy were evaluated by replicate dissolution analyses of K1919, an 86-yr-old Kilauea basalt (same lava as the more widely distributed BHVO; Supplementary Table 4). K1919 is nearly four ²¹⁰Pb half-lives old, extremely fresh, and compositionally similar to our unknowns. External precision for ²³⁸U, ²³⁰Th, ²²⁶Ra and ²¹⁰Pb ranged from 0.3% to 0.8%. Perhaps most pertinent is the quality of (²¹⁰Pb/²²⁶Ra) data because disequilibria are relatively small in our unknowns. External reproducibility was 0.3% (2σ) on four (²¹⁰Pb/²²⁶Ra) replicates. The mean (²¹⁰Pb/²²⁶Ra) is in secular equilibrium within

error (0.995 ± 0.008). Individual ratio analyses are $\leq 0.8\%$ of the secular equilibrium ratio, well inside the 1.5% combined abundance limitations of our ^{209}Po spike (a dilution of SRM 4326, which is abundance certified to 0.42%) and the widely used NIST SRM 4966 with which we calibrate our Ra spikes (abundance certified to 1.19%).

Data comparison with ref. 12 provides an additional independent assessment of our sample preparation and ^{238}U – ^{230}Th – ^{226}Ra analysis protocols because northern-EPR specimens 2504-1, 2372-1 and 2392-1 were analysed in both laboratories. The respective differences between reported ($^{230}\text{Th}/^{238}\text{U}$) and ($^{229}\text{Ra}/^{230}\text{Th}$) are 0.3% and -0.7% , 0.1% and 0.9%, and -0.8% and -1.9% (Supplementary Table 3). These are our lowest U-series abundance samples (typically the most difficult to pick, prepare and analyse), so this is a realistic ^{238}U – ^{230}Th – ^{226}Ra reproducibility upper limit.

Melt modelling. We used a commonly applied analytically solvable model² and upper melt column mineralogy (clinopyroxene–olivine–orthopyroxene–spinel) to investigate ^{210}Pb – ^{226}Ra – ^{230}Th – ^{238}U partitioning during mantle melting and melt transport. Large $D_{\text{Pb}} - D_{\text{Ra}}$ and $D_{\text{Ra}} - D_{\text{Th}}$ differences¹⁵ favour such partitioning and produce ^{210}Pb deficits and negatively correlated ($^{210}\text{Pb}/^{226}\text{Ra}$) versus ($^{226}\text{Ra}/^{230}\text{Th}$) trajectories over a broad range of melting rate (M in $\text{kg m}^{-3} \text{yr}^{-1}$) and porosity (ϕ) conditions. Garnet melting does not change curves much because pyroxene dominates Pb–Ra–Th partitioning. Conditions bound sample ($^{226}\text{Ra}/^{230}\text{Th}$) and ($^{230}\text{Th}/^{238}\text{U}$) in Fig. 3 examples. The similarity of Fig. 3a and b trajectories demonstrates ^{210}Pb – ^{226}Ra disequilibria insensitivity to M and D_i (ref. 15): (for clinopyroxene, $D_{\text{Ba}}(0.0005\text{--}0.005, D_{\text{Ba}}/D_{\text{Th}} \approx 0.4, D_{\text{Ba}}/D_{\text{Pb}} \approx 0.007\text{--}0.04, D_{\text{Ra}}/D_{\text{Ba}} \approx 0.1\text{--}0.07$; for olivine and orthopyroxene, D_{Ba} is much lower at ratios similar to $D_{\text{Pb}}, D_{\text{Ra}}, D_{\text{Th}}$).

Received 12 January; accepted 1 July 2005.

- Crisp, J. A. Rates of magma emplacement and volcanic output. *J. Volcanol. Geotherm. Res.* **20**, 177–211 (1984).
- Williams, R. W. & Gill, J. B. Effects of partial melting on the uranium decay series. *Geochim. Cosmochim. Acta* **53**, 1607–1619 (1989).
- Oversby, V. M. & Gast, P. W. Lead isotope compositions and uranium decay series disequilibrium in recent volcanic rocks. *Earth Planet. Sci. Lett.* **5**, 199–206 (1968).
- Rubin, K. H. & Macdougall, J. D. Submarine magma degassing and explosive magmatism at Macdonald (Tamarii) seamount. *Nature* **341**, 50–52 (1989).
- Rubin, K. H. *et al.* ^{238}U decay series systematics of young lavas from Batur volcano, Sunda Arc. *J. Volcanol. Geotherm. Res.* **38**, 215–226 (1989).
- Gill, J. B. & Williams, R. W. Th isotope and U-series studies of subduction-related volcanic rocks. *Geochim. Cosmochim. Acta* **54**, 1427–1442 (1990).
- Sigmarrsson, O. Short magma chamber residence time at an Icelandic volcano inferred from U-series disequilibria. *Nature* **382**, 440–441 (1996).
- Gauthier, P.-J. & Condomines, M. ^{210}Pb – ^{226}Ra radioactive disequilibria in recent lavas and radon degassing: inferences on the magma chamber dynamics at Stromboli and Merapi volcanoes. *Earth Planet. Sci. Lett.* **172**, 111–126 (1999).
- Turner, S., Black, S. & Berlo, K. ^{210}Pb – ^{226}Ra and ^{228}Ra – ^{232}Th systematics in young arc lavas: implications for magma degassing and ascent rates. *Earth Planet. Sci. Lett.* **227**, 1–16 (2004).
- Rubin, K. H. Degassing of metals and metalloids from erupting seamount and mid-ocean ridge volcanoes: Observations and predictions. *Geochim. Cosmochim. Acta* **61**, 3525–3542 (1997).
- Rubin, K. H. & Macdougall, J. D. ^{226}Ra excesses in mid-ocean-ridge basalts and mantle melting. *Nature* **335**, 158–161 (1988).
- Sims, K. W. W. *et al.* Chemical and isotopic constraints on the generation and transport of magma beneath the East Pacific Rise. *Geochim. Cosmochim. Acta* **66**, 3481–3504 (2002).
- Saal, A. E. & Van Orman, J. A. The ^{226}Ra enrichment in oceanic basalts: Evidence for melt-cumulate diffusive interaction processes within the oceanic lithosphere. *Geochim. Geophys. Geosyst.* **5**, Q02008, doi:10.1029/2003GC000620 (2004).
- Cherniak, D. J. Pb diffusion in clinopyroxene. *Chem. Geol.* **150**, 105–117 (1998).
- Blundy, J. D., Wood, B. J. *et al.* Mineral-melt partitioning of uranium, thorium and their daughters. in *Uranium-Series Geochemistry Rev. Mineral. Geochem.* (eds Bourdon, B. *et al.*) **52**, 59–118 (Mineral. Soc. Am., Washington DC, 2003).
- McKenzie, D. Constraints on melt generation and transport from U-series activity ratios. *Chem. Geol.* **162**, 81–94 (2000).
- Jull, M., Kelemen, P. B. & Sims, K. W. Consequences of diffuse and channeled porous melt migration on Uranium series. *Geochim. Cosmochim. Acta* **66**, 4133–4148 (2002).
- Rubin, K. H. *et al.* Magmatic history and volcanological insights from individual lava flows erupted on the seafloor. *Earth Planet. Sci. Lett.* **188**, 349–367 (2001).
- Perfit, M. R. & Chadwick, W. W. Jr. in *Faulting and Magmatism at Mid-Ocean Ridges* (eds Buck, W. R., Delaney, P. T., Karson, J. A. & Lagabriele, Y.) 59–115 (Geophys. Monogr. 106, Am. Geophys. Union, Washington DC, 1998).
- Sinton, J. *et al.* Volcanic eruptions on mid-ocean ridges: New evidence from the superfast-spreading East Pacific Rise, 17° – 19° S. *J. Geophys. Res.* **107**, doi:10.1029/2000JB000090 (2002).
- West, M., Menke, W., Tolstoy, M., Webb, S. & Sohn, R. Magma storage beneath Axial volcano on the Juan de Fuca mid-ocean ridge. *Nature* **413**, 833–836 (2001).
- Sims, K. W., DePaolo, D. J., Murrell, M. T., Baldrige, W. S. & Goldstein, S. J. Porosity of the melt zone and variations in solid mantle upwelling rates beneath Hawaii: inferences from ^{238}U – ^{230}Th – ^{226}Ra and ^{235}U – ^{231}Pa disequilibria. *Geochim. Cosmochim. Acta* **63**, 4119–4138 (1999).
- Pietruszka, A. J., Rubin, K. H. & Garcia, M. O. ^{226}Ra – ^{230}Th – ^{238}U disequilibria of historical Kilauea lavas (1790–1982) and the dynamics of mantle melting within the Hawaiian plume. *Earth Planet. Sci. Lett.* **186**, 15–31 (2001).
- Garcia, M. O. *et al.* Petrology and geochronology of basalt breccia from the 1996 earthquake swarm of Loihi seamount, Hawaii: magmatic history of its 1996 eruption. *Bull. Volcanol.* **59**, 577–592 (1998).
- Condomines, M., Morand, P. & Allegre, C. J. ^{230}Th – ^{238}U radioactive disequilibria in tholeiites from the FAMOUS zone (Mid-Atlantic Ridge, 36 deg 50 min N): Th and Sr isotopic geochemistry. *Earth Planet. Sci. Lett.* **55**, 247–256 (1981).
- Chadwick, D. J. *et al.* Magmatic effects of the Cobb hotspot on the Juan de Fuca ridge. *J. Geophys. Res.* **101**, B03101, doi:10.1029/2003JB002767 (2005).
- Rubin, K. H. Analysis of $^{232}\text{Th}/^{230}\text{Th}$ in volcanic rocks: a comparison of thermal ionization mass spectrometry and other methodologies. *Chem. Geol.* **175**, 755–782 (2001).
- Rubin, K. H., Macdougall, J. D. & Perfit, M. R. ^{210}Po – ^{210}Pb dating of recent volcanic eruptions on the sea floor. *Nature* **368**, 841–844 (1994).
- Smith, M. C., Perfit, M. R. & Jonasson, I. R. Petrology and geochemistry of basalts from the southern Juan de Fuca Ridge: controls on the spatial and temporal evolution of mid-ocean ridge basalt. *J. Geophys. Res.* **99**, 4787–4812 (1994).
- Embley, R. W., Chadwick, W. W. Jr, Clague, D. A., Stakes, D. *et al.* 1998 eruption of Axial Volcano; multibeam anomalies and seafloor observations. *Geophys. Res. Lett.* **26**, 3425–3428 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Perfit and J. Sinton for providing most of the samples, associated geological insight and unpublished trace element data used in this research, L. Sacks for performing some of the chemical analyses, D. Pyle and K. Spencer for effective management of the SOEST Isotope lab, K. Sims and T. Elliott for thorough reviews, and J. D. Macdougall on the occasion of his retirement for earlier work inspiring this effort. The US National Science Foundation supported this work.

Author Contributions K.H.R. designed the experiment; all authors participated in data acquisition and analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.R. (krubin@hawaii.edu).

Variations in earthquake-size distribution across different stress regimes

Danijel Schorlemmer¹, Stefan Wiemer¹ & Max Wyss²

The earthquake size distribution follows, in most instances, a power law^{1,2}, with the slope of this power law, the '*b* value', commonly used to describe the relative occurrence of large and small events (a high *b* value indicates a larger proportion of small earthquakes, and vice versa). Statistically significant variations of *b* values have been measured in laboratory experiments, mines and various tectonic regimes such as subducting slabs, near magma chambers, along fault zones and in aftershock zones³. However, it has remained uncertain whether these differences are due to differing stress regimes, as it was questionable that samples in small volumes (such as in laboratory specimens, mines and the shallow Earth's crust) are representative of earthquakes in general. Given the lack of physical understanding of these differences, the observation that *b* values approach the constant 1 if large volumes are sampled⁴ was interpreted to indicate that *b* = 1 is a universal constant for earthquakes in general⁵. Here we show that the *b* value varies systematically for different styles of faulting. We find that normal faulting events have the highest *b* values, thrust events the lowest and strike-slip events intermediate values. Given that thrust faults tend to be under higher stress than normal faults we infer that the *b* value acts as a stress meter that depends inversely on differential stress.

To resolve the issue of influence of stress on *b*, we search several high-quality large data sets, including a global one, for systematic dependence of *b* on the style of faulting. To estimate *b* values, we use the maximum-likelihood method⁶, with uncertainties estimated as described in ref. 7. We require at least 100 events per sample for computations of *b*, but mark all *b* values that are computed from less than 200 events. For consistency in the analysis we binned magnitudes in all catalogues to $\Delta M = 0.1$. In addition, we selected periods of homogeneous recording, determined their overall magnitude of completeness M_c for each catalogue and cut the catalogues accordingly (Table 1). The source depths were restricted to 50 km, because deep events may obey different laws of faulting. The significance of differences between *b* values is computed with the Utsu-test⁸ as the probability P_b that the *b* values are not different. Values of $\log P_b \leq -1.3$ and $\log P_b \leq -1.9$ indicate significant and highly significant differences, respectively.

Earthquake focal mechanisms are now determined routinely for thousands of events. For this study we use the best sources of focal mechanisms available worldwide: the global Harvard CMT moment tensor catalogue, a catalogue of relocated events of southern California (SCSN)⁹, a northern California data set (NCSN), a data set of events from the Kanto-Tokai region of Japan, and the F-Net data for all of Japan. From the Kanto-Tokai catalogue we excluded volcanic and geothermal areas (140.30° E, 36.99° N; 137.73° E, 35.71° N; 136.78° E, 35.74° N; 136.64° E, 35.25° N; 137.01° E, 34.58° N; 138.54° E, 34.58° N; 139.00° E, 34.78° N; 138.94° E, 35.34° N;

139.03° E, 35.34° N; 139.46° E, 34.82° N; 140.05° E, 34.91° N; 140.90° E, 35.69° N; 141.36° E, 36.57° N; 141.37° E, 36.72° N; 140.55° E, 36.98° N; 140.30° E, 36.99° N).

Two quality descriptors in the Californian catalogues (one in the NEID Kanto-Tokai catalogue, namely misfit) were used to limit the analysis to the high-quality events: solution misfit ≤ 0.2 and station distribution ratio ≥ 0.5 .

For classifying events as strike-slip, thrust or normal events, we use the rake angle λ within a given range of $\pm\gamma$ (Aki-Richards convention: $\lambda = 0$ or $\lambda = \pm 180^\circ$ as strike-slip, $\lambda = 90^\circ$ as thrust, and $\lambda = -90^\circ$ as normal events). Because the Californian catalogues contain only one nodal plane, we computed the corresponding second planes.

We performed two different computations. First, we computed *b* values as a function of the rake angles λ of the single planes with a constant range γ (b_λ plot). We used $\gamma = 20^\circ$ for the SCSN data, $\gamma = 60^\circ$ for the NEID F-Net data, and $\gamma = 40^\circ$ for the other catalogues. The different ranges were used to smooth the results according to the amount of available data. Second, to explore the separation of *b* values into different classes, we computed *b* values as a function of the range γ for all three classes of events (b_γ plot). Hereby, the γ value was varied from 175° to 5° in steps of -10° , and both rake angles λ of the two nodal planes had to fall within the given range. For $\gamma \geq 45^\circ$, events may fall into two or more classes.

We computed the b_λ plot for the SCSN data (Fig. 1a). This catalogue is probably the highest-quality regional data set of focal mechanisms available worldwide, because of the high density of stations, the relocation of events and recomputation of take-off angles with a three-dimensional velocity model. These data show that the differences in *b* values for the three classes of events are larger than the errors of individual samples and are significant (Table 2). Normal events (green) show the highest *b* values ($b_{NR} \approx 1.1$), strike-slip events (red) show intermediate values ($b_{SS} \approx 0.9$) and thrust events (blue) the lowest ($b_{TH} \approx 0.7$). The transition between the average *b* value and the extreme *b* values (normal, thrust) is pronounced.

In the b_γ plot of the SCSN data (Fig. 1b), one can see that with smaller γ , *b* values of normal and thrust events start to deviate

Table 1 | Selection parameters of the catalogues

Network	Period	M_c	Depth (km)	No. of events
Harvard CMT	1.1.1980–31.12.2004	5.5	0–50	7,636
SCSN	1.1.1981–31.12.2003	2.5	all	14,003
NCSN	1.1.1981–31.12.2004	3.0	all	4,250
NEID F-Net	4.1.1997–31.12.2004	4.5	0–50	1,579
NEID Kanto-Tokai	1.1.1982–1.7.2003	3.0	0–50	2,337

¹Swiss Seismological Service, ETH Zürich, ETH Hönggerberg, Schafmattstr. 30, 8093 Zürich, Switzerland. ²World Agency of Planetary Monitoring and Earthquake Risk Reduction, Route de Jargonnant 2, 1207 Genève, Switzerland.

systematically from the regional average and reach their extreme for pure event classes with range $\gamma \leq 5^\circ$. Strike-slip events show only minor deviation from the regional average b value, regardless of the range γ . Only for $\gamma = 5^\circ$ is the b value significantly higher than the average. The differences in the frequency–magnitude distribution between pure thrust and pure normal events are statistically highly significant (Fig. 1c and Table 2).

The b -value variations of the Harvard catalogue are less pronounced; however, they show the same dependence on faulting style (Fig. 1d, e). The separation is clearer when using only the first nodal plane for classifying events (Fig. 1e, inset) rather than both nodal planes (Fig. 1e). Although each catalogue covers a different range of magnitudes, the frequency–magnitude distributions emphasize the similarity of the b values for the two displayed classes of events (Fig. 1c).

The data in all three additional catalogues show essentially the same systematic dependence of b on faulting style (Fig. 2). The absolute level of b does not concern us here, because it can vary in different regions and for different networks. The important observation is that, in all cases, normal events have the highest b values, followed by strike-slip and thrust events. In some data sets the separation is no longer clear for the narrowest ranges γ of λ because the error bars increase as a result of small sample sizes.

The data sets analysed are very different in depth distribution, magnitude range and tectonic environment. Nevertheless, all data sets show the same dependence of b on focal mechanisms.

The uniformity of the separation of b values according to faulting type in these tectonically different data sets requires a universal interpretation. We propose that the differential stress, and indirectly the confining pressure (to which the differential stress is tied), is the parameter most strongly controlling faulting types, thus influencing differences in b . This inverse relationship follows because we show that $b_{TH} < b_{SS} < b_{NR}$, and it is known that for a given vertical stress σ_v the mean stresses obey the relationship $\bar{\sigma}_{TH} > \bar{\sigma}_{SS} > \bar{\sigma}_{NR}$.

Our results imply that the inverse relationship between differential stress $\Delta\sigma$ and b is universally valid, spanning the range from submillimetres to hundreds of kilometres of rupture length. The b value can therefore be interpreted as a ‘stressmeter’ in the Earth’s crust. However, locally, secondary effects can come from material properties^{10,11} and from modifications of the stress tensor by pore pressure¹².

The idea that depths might be the controlling factor, in general, can be ruled out because the average depths of the three classes of events are not systematically ordered according to the b values. Depth is also not a good candidate to be a universal factor controlling b , because the trends reported are opposite in California^{13,14} and Japan¹⁵. The degree of material heterogeneity¹⁰ as a fundamental cause of these universal differences can also be ruled out, because the observation is valid for small as well as larger earthquakes and for all existing tectonic settings.

Previous studies^{4,16} based on fewer data and a less quantitative analysis of focal mechanism reported that the size distribution β of moments in the Harvard CMT catalogue (0–70 km depth) obey the relationship $\beta_{NR} > \beta_{TH} > \beta_{SS}$. However, the most comparable computation of Kagan, using the entire catalogue in the period 1987–1999 with a completeness of $M_c = 5.4$ reproduces the relation that we found: $(\beta_{NR} = 0.731) > (\beta_{SS} = 0.643) > (\beta_{TH} = 0.642)$.

The inverse relationship between stress and b is consistent with a range of other observations: laboratory rock specimens^{11,17}, mines¹⁸ and increased pore pressure^{12,19} (which decreases the differential stress). For the Los Angeles Basin, it was found²⁰ that $b_{SS} = 1.1$ and $b_{TH} = 0.7$ for strike-slip and thrust events, respectively. Although this result was based on only 144 and 78 earthquakes, respectively, it supports our findings. Finally, the strong differences in b between locked parts of faults, where b values are low, and creeping sections with much higher b (refs 21–23), are consistent with the model. The shear strength is assumed to be higher in a locked section than in creeping sections. This results in high $\Delta\sigma$ and low b values for

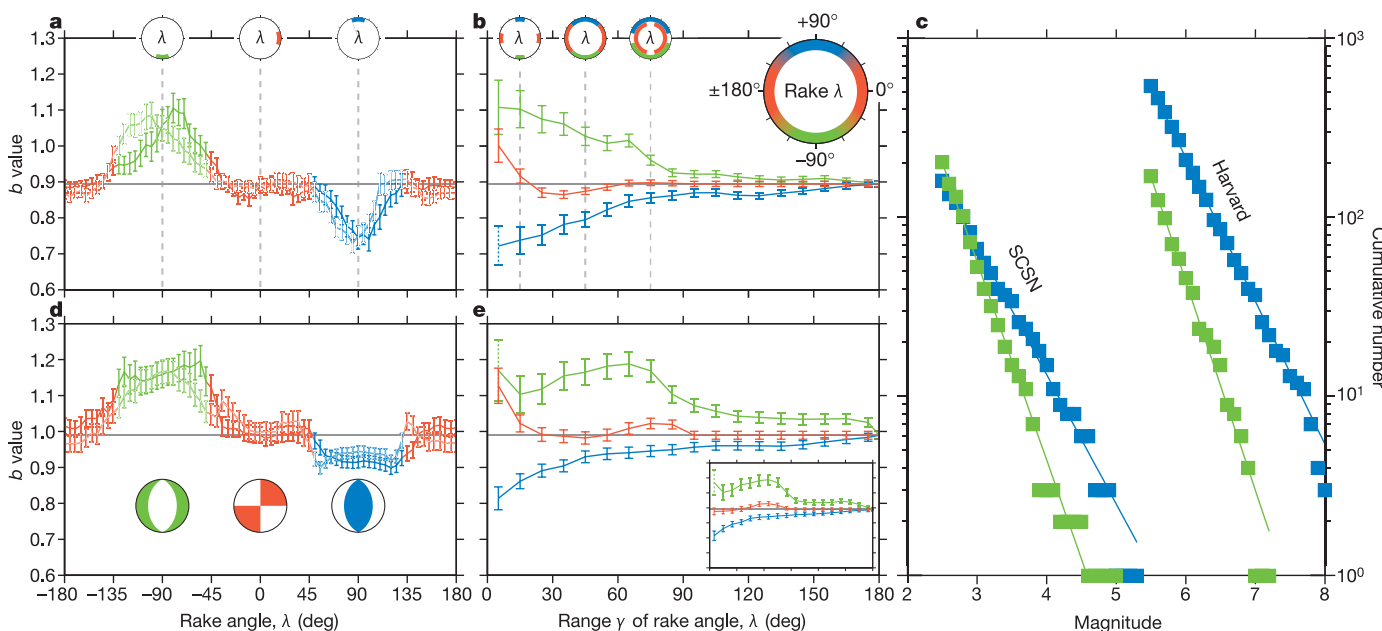


Figure 1 | Plots of b against rake angle and range of rake angle. **a**, b_λ plot of events in southern California, from the SCSN catalogue. In all frames the green, red and blue lines (solid, first plane; outlined, second plane) mark the b values of mainly normal, strike-slip and thrust events, respectively; the grey line marks the average b value and the vertical bars indicate the standard error⁷. (Solid bars are used for samples with $N \geq 200$, and dashed bars for samples with $200 > N \geq 100$). The circles at the top of the frame show the rake λ used for computing the b values $\lambda = -90^\circ \pm \gamma$, $\lambda = 0^\circ \pm \gamma$

and $\lambda = 90^\circ \pm \gamma$, $\gamma = 20^\circ$. **b**, b_γ plot of events in southern California. The circles at the top of the frame show the range of rake λ used for computing the b values ($\gamma = 15^\circ, 45^\circ$ and 75°). Inset, circle explaining the rake values and the corresponding colours of the classes of events. **c**, Frequency–magnitude distributions for pure normal (green) and pure thrust (blue) events of the SCSN and Harvard catalogues ($\gamma = 5^\circ$). **d**, As **a** for the Harvard catalogue. **e**, As **b** for the Harvard catalogue. Inset, as main panel but considering only the rake of the first nodal plane.

Table 2 | Values of b for three different ranges γ of the three classes of events and their Utsu-test probabilities ($\log P_b$) of being different

Network	Range γ (deg)	b_{TH}	b_{SS}	b_{NR}	$\log P_b$ TH-SS	$\log P_b$ SS-NR	$\log P_b$ TH-NR
SCSN	05	0.72 ± 0.05	1.00 ± 0.05	1.11 ± 0.08	-2.83	-0.30	-3.61
	25	0.75 ± 0.03	0.87 ± 0.01	1.07 ± 0.04	-2.52	-5.95	-9.49
	45	0.79 ± 0.02	0.87 ± 0.01	1.03 ± 0.03	-2.32	-7.17	-10.38
Harvard	05	0.81 ± 0.03	1.13 ± 0.05	1.17 ± 0.08	-5.19	-0.04	-3.67
	25	0.89 ± 0.02	0.99 ± 0.02	1.12 ± 0.04	-2.67	-1.55	-5.52
	45	0.93 ± 0.02	0.98 ± 0.02	1.17 ± 0.04	-0.94	-4.51	-8.07
NCSN	05	-	-	-	-	-	-
	25	0.67 ± 0.05	0.81 ± 0.02	1.01 ± 0.06	-1.14	-1.99	-3.61
	45	0.73 ± 0.04	0.91 ± 0.02	1.00 ± 0.04	-3.50	-0.74	-4.52
Kanto-Tokai	05	-	-	-	-	-	-
	25	0.68 ± 0.03	0.89 ± 0.05	1.03 ± 0.09	-2.72	-0.35	-2.78
	45	0.66 ± 0.02	0.90 ± 0.04	1.06 ± 0.07	-7.50	-0.79	-6.44
F-Net	05	-	-	-	-	-	-
	25	0.80 ± 0.04	1.02 ± 0.07	1.06 ± 0.08	-1.50	-0.02	-1.84
	45	0.78 ± 0.03	1.09 ± 0.05	1.05 ± 0.05	-5.28	-0.05	-3.57

Significant $\log P_b$ values are bold.

locked sections, whereas creeping sections exhibit low $\Delta\sigma$ and high b values. This conceptual model relating b to fault locking was recently tested and supported by the magnitude 6.0 Parkfield, California, event in 2004 (ref. 24), which almost exclusively ruptured areas of the San Andreas fault previously mapped as regions of low b .

The physical model offered by Scholz¹¹ as an explanation for the stress dependence of b emphasizes the role of differential stress. He argues that in a rock mass with varying resistance to faulting (and varying stress concentration) a rupture is likely to connect one high-stress subvolume to the next, because the system contains more

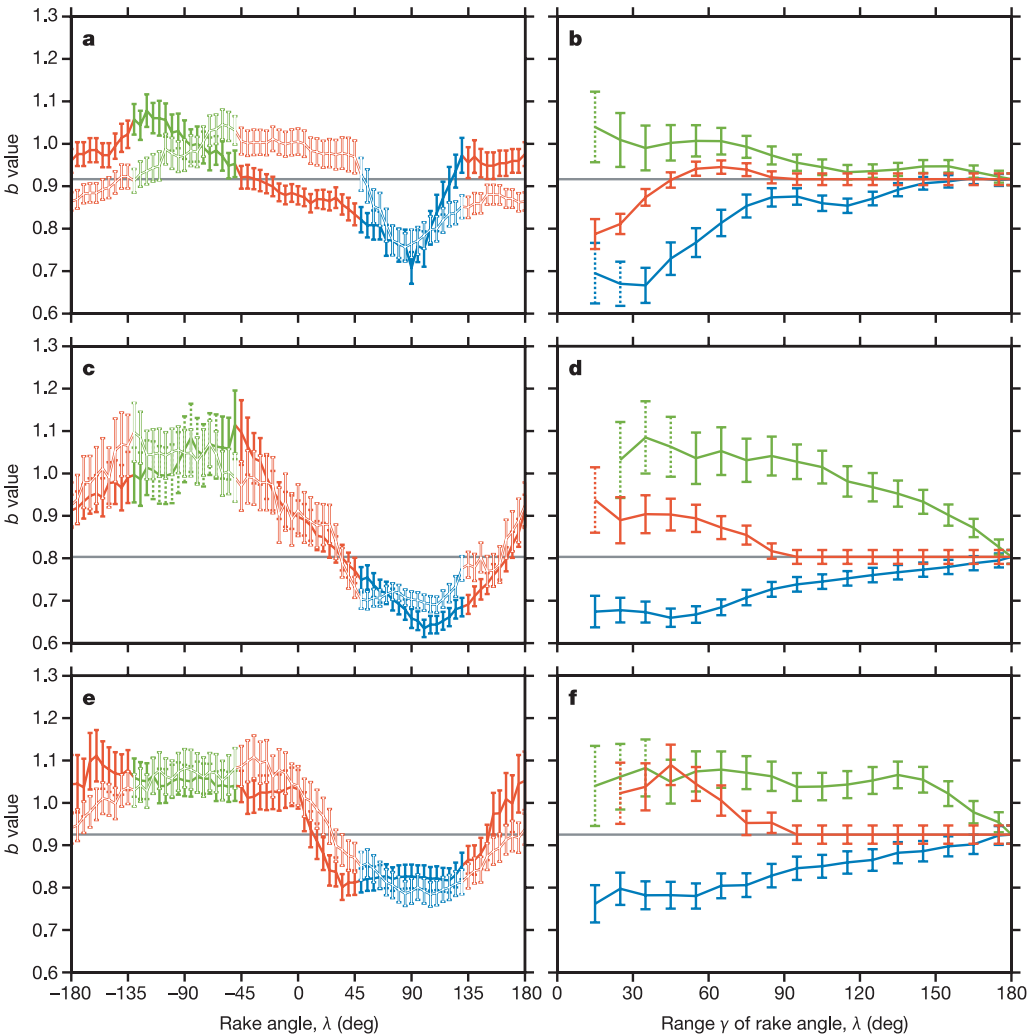


Figure 2 | b_λ plots and b_γ plots for the NCSN catalogue (a, b), the NEID Kanto-Tokai catalogue (c, d) and the NEID F-Net catalogue (e, f). In all frames the green, red and blue lines (solid, first plane; outlined, second plane) mark the b values of mainly normal, strike-slip and thrust events,

respectively; the grey line marks the average b value and the vertical bars indicate the standard error⁷. (Solid bars are used for samples with $N \geq 200$, dashed bars for samples with $200 > N \geq 100$).

energy. Thus, earthquakes grow larger in a high-stress environment. This finding is additionally supported by fractal simulations²⁵. Amitrano¹⁷ proposed a more geometrical model, emphasizing the role of confining pressure. On the basis of a numerical model, he argued that, at low confining pressure, the internal friction angle is such that failure in a given element can influence other elements only in restricted directions. At high confining pressure, the angle of internal friction decreases and interactions with elements containing defects are possible in all directions. Thus, failures grow larger in a system under high confining pressure.

The dependence of b on differential stress suggests that a systematic decrease in b towards the end of the seismic cycle^{26,27} could indeed exist. However, the amplitude of the differential stress increase during the loading cycle is probably an order of magnitude smaller than in laboratory experiments and in mines. The high-quality data sets necessary to test this hypothesis are probably not yet available.

One of the implications of the universal dependence of b on differential stress is that earthquake hazard analysis should be rethought. The b value is a critical parameter in studies of seismic hazard. If events of different faulting styles within one seismogenic zone follow different recurrence laws, and because faulting styles influence ground motions²⁸, estimating the recurrence for each faulting style separately may produce more accurate estimates of seismic hazard. In addition, hazard studies should pay attention to spatial heterogeneity in b values, because it is directly related to the distribution of differential stresses in the Earth's crust.

Received 1 February; accepted 2 August 2005.

- Ishimoto, M. & Iida, K. Observations of earthquakes registered with the microseismograph constructed recently. *Bull. Earthquake Res. Inst. Tokyo Univ.* **17**, 443–478 (1939).
- Gutenberg, B. & Richter, C. F. Frequency of earthquakes in California. *Bull. Seismol. Soc. Am.* **34**, 185–188 (1944).
- Wiemer, S. & Wyss, M. Mapping spatial variability of the frequency–magnitude distribution of earthquakes. *Adv. Geophys.* **45**, 259–302 (2002).
- Frohlich, C. & Davis, S. D. Teleseismic b values; or, much ado about 1.0. *J. Geophys. Res.* **98**, 631–644 (1993).
- Kagan, Y. Y. Universality of the seismic moment–frequency relation. *Pure Appl. Geophys.* **155**, 537–574 (1999).
- Bender, B. Maximum likelihood estimation of b values for magnitude grouped data. *Bull. Seismol. Soc. Am.* **73**, 831–851 (1983).
- Shi, Y. & Bolt, B. A. The standard error of the magnitude–frequency b -value. *Bull. Seismol. Soc. Am.* **72**, 1677–1687 (1982).
- Utsu, T. *Report of the Joint Research Institute for Statistical Mathematics* Vol. 34, 139–157 (Institute for Statistical Mathematics, Tokyo, 1992).
- Hauksson, E. Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California. *J. Geophys. Res.* **105**, 13875–13903 (2000).
- Mogi, K. Magnitude–frequency relations for elastic shocks accompanying fractures of various materials and some related problems in earthquakes. *Bull. Earthquake Res. Inst. Univ. Tokyo* **40**, 831–853 (1962).
- Scholz, C. H. The frequency–magnitude relation of microfracturing in rock and its relation to earthquakes. *Bull. Seismol. Soc. Am.* **58**, 399–415 (1968).
- Wyss, M. Towards a physical understanding of the earthquake frequency distribution. *Geophys. J. R. Astron. Soc.* **31**, 341–359 (1973).
- Mori, J. & Abercrombie, R. E. Depth dependence of earthquake frequency–magnitude distributions in California: Implications for rupture initiation. *J. Geophys. Res.* **102**, 15081–15090 (1997).
- Gerstenberger, M., Wiemer, S. & Giardini, D. A systematic test of the hypothesis that the b value varies with depth in California. *Geophys. Res. Lett.* **28**, 57–60 (2001).
- Wyss, M. & Matsumura, S. Most likely locations of large earthquakes in the Kanto and Tokai areas, Japan, based on the local recurrence times. *Physics of the Earth and Planetary Interiors* **131**, 173–184 (2002).
- Kagan, Y. Y. Seismic moment distribution revisited: I. Statistical results. *Geophys. J. Int.* **148**, 520–541 (2002).
- Amitrano, D. Brittle–ductile transition and associated seismicity: Experimental and numerical studies and relationship with the b value. *J. Geophys. Res.*, B2044 (2003) (doi:10.1029/2001JB000680).
- Urbancic, T. I., Trifu, C. I., Long, J. M. & Young, R. P. Space-time correlations of b -values with stress release. *Pure Appl. Geophys.* **139**, 449–462 (1992).
- Wiemer, S., McNutt, S. R. & Wyss, M. Temporal and three-dimensional spatial analysis of the frequency–magnitude distribution near Long Valley caldera. *California. Geophys. J. Int.* **134**, 409–421 (1998).
- Hauksson, E. Earthquakes, faulting, and stress in the Los Angeles basin. *J. Geophys. Res.* **95**, 15365–15394 (1990).
- Amelung, F. & King, G. The difference between earthquake scaling laws for creeping and non-creeping faults. *Geophys. Res. Lett.* **24**, 507–510 (1997).
- Wiemer, S. & Wyss, M. Mapping the frequency–magnitude distribution in asperities: An improved technique to calculate recurrence times? *J. Geophys. Res.* **102**, 15115–15128 (1997).
- Schorlemmer, D., Wiemer, S. & Wyss, M. Earthquake statistics at Parkfield: 1. Stationarity of b -values. *J. Geophys. Res.* **109**, B12307 (2004) (doi:10.1029/2004JB003234).
- Schorlemmer, D. & Wiemer, S. Microseismicity data forecast rupture area. *Nature* **434**, 1086 (2005).
- Huang, J. & Turcotte, D. L. Fractal distributions of stress and strength and variations of b -value. *Earth Planet. Sci. Lett.* **91**, 223–230 (1988).
- Smith, W. D. The b -values as an earthquake precursor. *Nature* **289**, 136–139 (1981).
- Lei, X. *et al.* Detailed analysis of acoustic emission activity during catastrophic fracture of faults in rock. *J. Struct. Geol.* **26**, 247–258 (2004).
- Oglesby, D. D., Archuleta, R. J. & Nielsen, S. B. Dynamics of dip-slip faulting: Explorations in two dimensions. *J. Geophys. Res.* **105**, 13643–13653 (2000).

Acknowledgements We thank E. Hauksson, D. Giardini, M. Mai, M. Gerstenberger, D. Jackson, J. Woessner and G. Hillers for discussions. We thank the Northern California Seismic Network, US Geological Survey, Menlo Park, and the Berkeley Seismological Laboratory, University of California, Berkeley, for the catalogue including phase data, and the National Research Institute for Earth Science and Disaster Prevention for mechanism solutions of the Kanto-Tokai area.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.S. (danijel@sed.ethz.ch).

Isolation of an autotrophic ammonia-oxidizing marine archaeon

Martin Könneke^{1*†}, Anne E. Bernhard^{1*†}, José R. de la Torre^{1*}, Christopher B. Walker¹, John B. Waterbury² & David A. Stahl¹

For years, microbiologists characterized the Archaea as obligate extremophiles that thrive in environments too harsh for other organisms. The limited physiological diversity among cultivated Archaea suggested that these organisms were metabolically constrained to a few environmental niches. For instance, all Crenarchaeota that are currently cultivated are sulphur-metabolizing thermophiles¹. However, landmark studies using cultivation-independent methods uncovered vast numbers of Crenarchaeota in cold oxic ocean waters^{2,3}. Subsequent molecular surveys demonstrated the ubiquity of these low-temperature Crenarchaeota in aquatic and terrestrial environments⁴. The numerical dominance of marine Crenarchaeota—estimated at 10^{28} cells in the world's oceans⁵—suggests that they have a major role in global biogeochemical cycles. Indeed, isotopic analyses of marine crenarchaeal lipids suggest that these planktonic Archaea fix inorganic carbon⁶. Here we report the isolation of a marine crenarchaeote that grows chemolithoautotrophically by aerobically oxidizing ammonia to nitrite—the first observation of nitrification in the Archaea. The autotrophic metabolism of this isolate, and its close phylogenetic relationship to environmental marine crenarchaeal sequences, suggests that nitrifying marine Crenarchaeota may be important to global carbon and nitrogen cycles.

Since their discovery by Fuhrman *et al.* and DeLong over a decade ago^{2,3}, marine Crenarchaeota are now recognized to be a dominant fraction of bacterioplankton in the ocean. These microorganisms can account for up to 40% of the bacterioplankton in deep ocean waters⁵. Although there are no known low-temperature Crenarchaeota in culture, compound-specific $\Delta^{14}\text{C}$ analysis of lipid biomarkers⁶ and studies of ^{13}C -bicarbonate tracer uptake by natural populations⁷ have suggested autotrophy. However, another study of natural populations demonstrated the uptake of tracer levels of tritiated amino acids, suggesting some use of fixed carbon⁸. We expect that the availability of a representative organism in pure culture will facilitate studies of their physiology and evolutionary origin, and help in understanding their contribution to oceanic biogeochemical cycles.

Known nitrifying bacteria fall into two distinct physiological groups: those that oxidize ammonia to nitrite, and others that oxidize nitrite to nitrate⁹. None has been shown to oxidize ammonia completely to nitrate. Existing genera of ammonia-oxidizing bacteria (AOB) fall within the Betaproteobacteria and the Gammaproteobacteria (ref. 9). Molecular studies of AOB in nitrifying systems, including aquaria¹⁰, have been limited to these two phylogenetic groups⁹. We first suspected an involvement of Archaea in nitrification after we completed several cultivation-independent ribosomal RNA gene surveys of nitrifying environments. We detected sequences

affiliated with the marine group 1 Crenarchaeota in nitrifying dilution cultures developed from Plum Island Sound (Massachusetts) estuary sediment, in nitrifying filtration systems at the Shedd Aquarium (Chicago, Illinois), and in gravel from a marine tropical fish tank at the Seattle Aquarium (Seattle, Washington).

Further evidence for archaeal nitrifiers resulted from ammonia-oxidizing cultures highly enriched in marine group 1 Crenarchaeota. Filtered aquarium water (0.2- μm polyethersulphone membrane; Nalgene) supplemented with 1 mM ammonium chloride was inoculated with gravel from a tropical marine tank at the Seattle Aquarium. Cultures enriched for Crenarchaeota were incubated at 21–23 °C in the dark. Repeated serial transfers of 10% of the culture volume into fresh aquarium-water medium resulted in an enrichment comprised of approximately 90% Crenarchaeota and 10% organisms affiliated with the bacterial domain after six months (data not shown). Characterization of this highly enriched culture revealed that oxidation rates of ammonia to nitrite corresponded with increasing abundance of Crenarchaeota (measured by quantitative polymerase chain reaction (PCR); Supplementary Information) indicating nitrification (data not shown).

After initial enrichment, the Crenarchaeota were isolated in a defined medium (see Methods) containing bicarbonate and ammonia as the sole carbon and energy sources, suggesting autotrophy. A pure culture of Crenarchaeota (designated SCM1) was recovered after three serial end-point dilutions in this medium, facilitated by the addition of streptomycin and filtration of the inoculum through a 0.45- μm HT Tuffryn membrane syringe filter (Pall). The purity of SCM1 was confirmed by quantitative PCR and fluorescent *in situ* hybridization (FISH), and supported by a failure to recover bacterial 16S rRNA genes by PCR amplification or to promote the growth of heterotrophic bacteria by the addition of yeast extract and peptone to the defined culture medium (data not shown). PCR amplification of nearly full-length 16S rRNA genes from SCM1 identified only crenarchaeal sequences. The clonal structure of SCM1 was confirmed by comparing the sequences of PCR-amplified fragments of 1,650 base pairs (bp) in length and containing most of the 16S rRNA, the complete 16S–23S internal transcribed spacer, and a small portion of the 23S rRNA gene (Supplementary Information).

Comparative sequence analysis of 16S rRNA genes revealed a high level of sequence identity (>98%) between SCM1 and marine group 1 Crenarchaeota sequences recovered from the North Atlantic, the Red Sea, the Antarctic and hydrothermal vents (Fig. 1). Phylogenetic analysis indicates that all marine group 1 Crenarchaeota—including SCM1, crenarchaeal sequences from the Sargasso Sea¹¹ and *Cenarchaeum symbiosum* (an uncultured marine sponge symbiont)¹²—form a monophyletic clade sharing >94% rRNA sequence identity (Fig. 1).

¹Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington 98195, USA. ²Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [†]Present addresses: Institut für Chemie und Biologie des Meeres, Universität Oldenburg, Oldenburg 26111, Germany (M.K.); Department of Biology, Connecticut College, New London, Connecticut 06320, USA (A.E.B.).

*These authors contributed equally to this work.

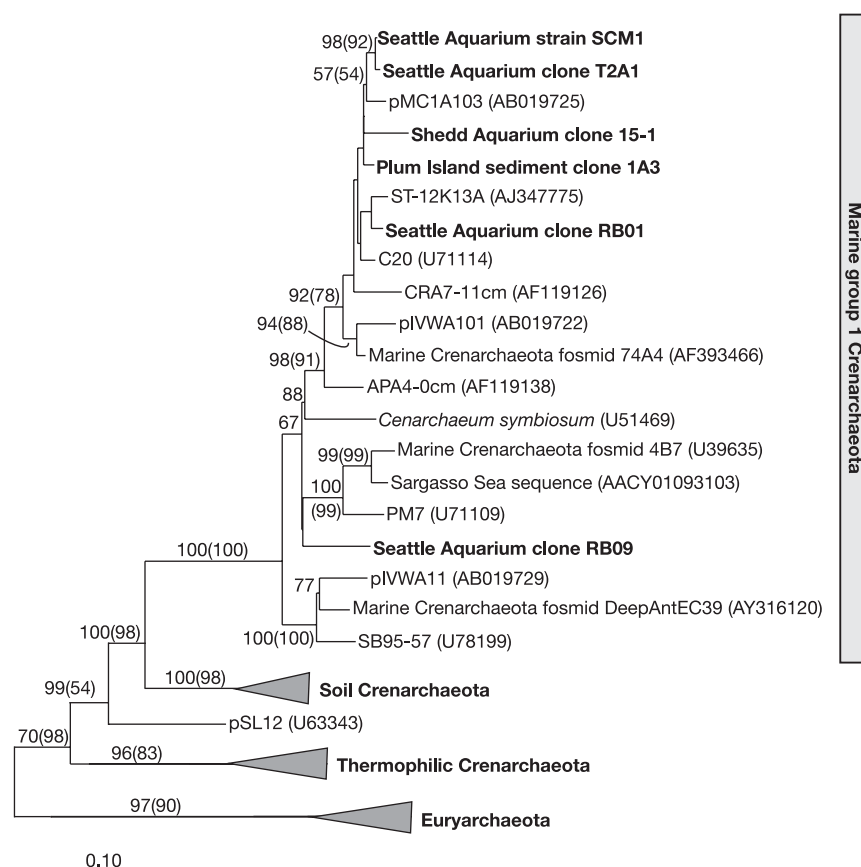


Figure 1 | Phylogenetic relationships between 16S rRNA sequences from SCM1 and representatives of the marine group 1 Crenarchaeota. The tree was constructed using the neighbour-joining algorithm with the Kimura two-parameter correction (1,218 positions). Nodes supported by bootstrap values >50% by neighbour-joining and parsimony (in parentheses) are

indicated. The sequences from the Shedd Aquarium and Plum Island sediment clones are short (674 and 720 bp, respectively) and were added to the tree using the parsimony tool in ARB (ref. 25). The scale bar represents 0.1 nucleotide changes per position.

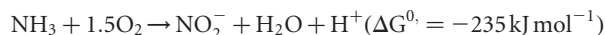
In contrast, members of the marine group 1 Crenarchaeota share only 84% 16S rRNA sequence identity with low-temperature Crenarchaeota found in soil, and less than 80% sequence identity with cultivated thermophilic Crenarchaeota. These differences in 16S rRNA sequences are consistent with variations found in genome fragments from marine and soil Crenarchaeota^{13,14}. Nevertheless, phylogenetic analyses clearly indicate that all low-temperature Crenarchaeota are more closely related to each other than to their thermophilic relatives (Fig. 1).

Cells of SCM1 visualized by electron microscopy appear as straight rods with a diameter of 0.17–0.22 μm and a length of 0.5–0.9 μm (Fig. 2). Cells occurred individually or in loose aggregates. Neither flagella nor intracellular compartments were apparent by electron microscopy. The size and morphology of SCM1 cells are nearly identical to marine Crenarchaeota in natural samples visualized by FISH^{12,15}. Individual SCM1 cells stained with 16S rRNA polyribonucleotide probes showed strong fluorescence at both poles and weak staining in the central region corresponding to the location of the nucleoid, identified by staining with the fluorescent DNA-binding dye 4',6'-diamidino-2-phenylindole (DAPI) (Fig. 2a and b; arrows), giving labelled cells the characteristic peanut-like shape previously reported for marine Crenarchaeota^{12,15}.

SCM1 grew to a maximal density of 1.4×10^7 cells ml^{-1} at 28 °C in defined medium containing 500 μM ammonium, with a minimum generation time of 21 h (Fig. 3). This cell density is approximately three orders of magnitude greater than that observed for marine Crenarchaeota in natural bacterioplankton samples³. Ammonium typically reaches concentrations of <0.03–1 μM in the open ocean and <0.03–100 μM in coastal waters¹⁶. Although this may ultimately

limit growth in the marine environment, it is possible that the marine Crenarchaeota are responsible for keeping ammonia concentrations low. The maximum growth rate of SCM1 in culture (0.78 d^{-1}) was somewhat higher than the range of rates estimated for natural bacterioplankton communities, which vary between 0.05 and 0.3 d^{-1} (ref. 17). Notably, the addition of organic compounds, even in very low concentrations, appeared to inhibit the growth of SCM1 (data not shown). Thus, organic material excreted by other organisms (for example, phototrophic primary producers) and a low concentration of ammonium may limit the abundance of marine Crenarchaeota in the environment.

Growth of SCM1 is correlated with near-stoichiometric conversion of ammonia to nitrite (Fig. 3). Among characterized AOB, ammonia monooxygenase (AMO) oxidizes ammonia to hydroxylamine, which is further oxidized to nitrite by hydroxylamine oxidoreductase⁹. The overall reaction is represented by the following stoichiometry:



Recently, AMO-related genes have been reported in environmental sequences of marine Crenarchaeota from the Sargasso Sea¹¹. Using comparisons of these genes to genome fragments from soil Crenarchaeota¹⁴ (Supplementary Data), we designed oligonucleotide primers to amplify and clone orthologues of the putative A, B and C subunits of AMO from SCM1. The predicted amino acid sequences of the putative AMO-encoding genes from SCM1 are very similar to sequences from Sargasso Sea and soil Crenarchaeota (93–98% and 80–90% amino acid sequence similarity, respectively; Supplementary Information), and are of low similarity to bacterial AMO-encoding

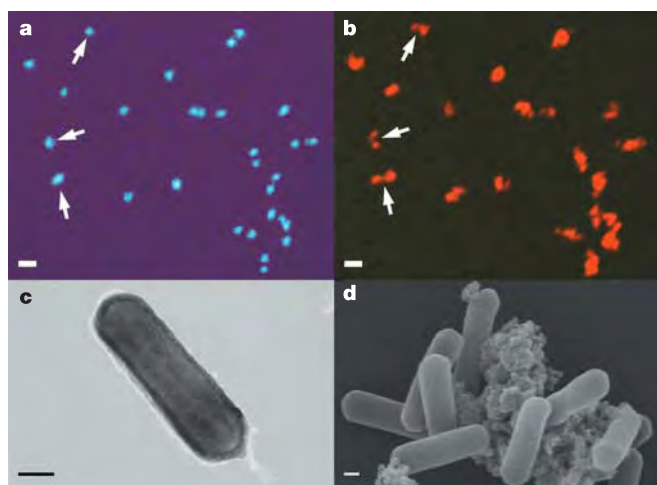


Figure 2 | Photomicrographs of SCM1. **a, b**, Fluorescence image of cells in identical fields of view stained with DAPI (**a**) and after hybridization with nucleotide polyprobes targeting SCM1 cells (**b**). Arrows indicate cells showing the characteristic peanut-like shape of marine Crenarchaeota^{12,15}. Scale bars represent 1 μm . **c**, Transmission electron micrograph of negative-stained cells. Scale bar represents 0.1 μm . **d**, Scanning electron micrograph of Au/Pd-sputtered cells. Scale bar represents 0.1 μm .

genes (38–51% amino acid sequence similarity). Although it remains uncertain whether these archaeal genes are orthologues of bacterial AMO-encoding genes, our studies correlate for the first time the presence of these archaeal AMO-encoding genes with nitrification. The presence of putative AMO-encoding genes in divergent marine and soil Crenarchaeota implies a broad distribution of nitrifying physiology in these organisms.

SCM1 oxidizes ammonia in the absence of organic carbon, conclusively demonstrating that SCM1 fixes inorganic carbon, a trait common to all known nitrifiers⁹. Isotopic analyses of membrane lipids extracted from the environment implicate the 3-hydroxypropionate carbon-fixation pathway in marine Crenarchaeota⁶. This pathway also exists in thermophilic Crenarchaeota such as *Acidianus infernus*, *Sulfolobus metallicus* and *Metallosphaera medulla*¹⁸. However, as other carbon fixation pathways also exist in these organisms¹⁹, the elucidation of the pathway(s) associated with mesophilic Crenarchaeota will require further investigation.

SCM1 is the first reported chemolithoautotrophic nitrifier in the domain Archaea and the first mesophilic isolate within the phylum Crenarchaeota. We propose the following candidate status:

Nitrosopumilales order nov.

Nitrosopumilaceae fam. nov.

'Nitrosopumilus maritimus' gen. et sp. nov.

Etymology. *nitrosus* (Latin masculine adjective): nitrous; *pumilus* (Latin masculine adjective): dwarf; *maritimus* (Latin masculine adjective): belonging to the sea. The name alludes to the habitat and size of the organism, in addition to its ability to convert ammonia to nitrite.

Locality. The rocky substratum of a tropical marine tank at the Seattle Aquarium (Seattle, Washington, USA).

Diagnosis. A chemolithoautotrophic nitrifier of the domain Archaea, appearing as straight rods with a diameter of 0.17–0.22 μm and a length of 0.5–0.9 μm .

Although we acknowledge that *'N. maritimus'* was not isolated directly from the open ocean, its close phylogenetic relationship to the marine group 1 Crenarchaeota, and the high amino acid sequence similarity between putative AMO-encoding genes from our isolate and from marine crenarchaeal environmental sequences¹¹, raises the possibility that nitrifying Crenarchaeota contribute to marine carbon and nitrogen cycles. Chemolithoautotrophy enables organisms such

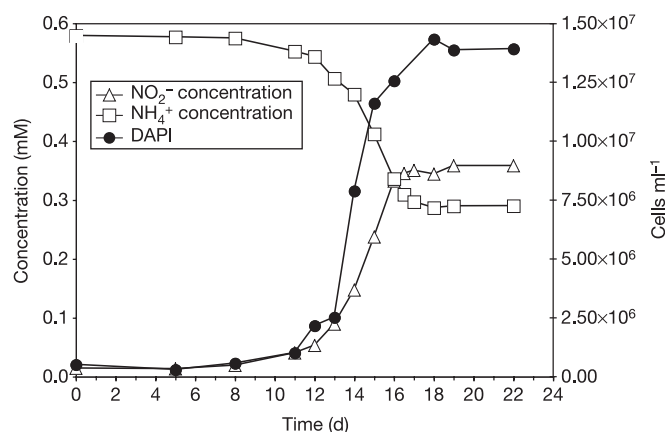


Figure 3 | Near-stoichiometric conversion of ammonia to nitrite by SCM1. Growth of SCM1 in Synthetic Crenarchaeota Media containing ammonium chloride and bicarbonate as sole energy and carbon sources, respectively. DAPI-stained cells were directly counted on filters by fluorescence microscopy. Ammonium consumption and nitrite production were determined in triplicate as described previously²⁷.

as *'N. maritimus'* to inhabit oligotrophic environments where they may function as important primary producers. In environments devoid of organic energy sources and sunlight, the oxidation of ammonia could contribute to primary productivity, and may explain the success of some marine Crenarchaeota in ecological niches such as the deep ocean and polar surface waters during winter⁴. This hypothesis is further supported by the parallels between the depth of the marine nitrite maximum²⁰ and the vertical distribution of Crenarchaeota in the ocean⁵.

Several authors have proposed that the low-temperature Crenarchaeota are derived from a thermophilic lineage and evolved to colonize mesophilic environments⁴. Previous studies of Yellowstone National Park (USA) hot springs have identified crenarchaeal 16S rRNA sequences that are phylogenetically most related to the low temperature Crenarchaeota²¹ (clone pSL12, Fig. 1). A recent study has reported the presence of crenarchaeol, a membrane lipid previously thought to be found exclusively in marine Crenarchaeota, in terrestrial hydrothermal springs²². High concentrations of bicarbonate were correlated with the abundance of crenarchaeol in these hydrothermal systems. These observations raise the possibility of the existence of thermophilic ammonia-oxidizing Crenarchaeota. Further biochemical and genomic studies of *'N. maritimus'* will provide a foundation for a more complete census of the habitat range of nitrifiers. The discovery of a nitrifying archaeon also highlights the questions of whether ammonia oxidation originated within the bacteria or the archaea, and whether the archetypical nitrifier was a thermophile.

METHODS

Cultivation. Cultures were grown aerobically at 28 °C in Synthetic Crenarchaeota Media containing NaCl (26 g l⁻¹), MgCl₂·6H₂O (5 g l⁻¹), MgSO₄·7H₂O (5 g l⁻¹), CaCl₂ (1.5 g l⁻¹) and KBr (0.1 g l⁻¹). After autoclaving, 1 ml non-chelated trace element mixture²³, 1 ml vitamin solution²³, 10 ml KH₂PO₄ solution (4 g l⁻¹), 1 ml selenite-tungstate solution²³, 1 ml bicarbonate solution (1 M), and 0.5–1 ml ammonium chloride (1 M) were added aseptically per litre of media. The pH was adjusted to 7.0–7.2 using NaOH. Growth was monitored by microscopy, quantitative PCR and nitrite production.

16S rRNA gene sequence and phylogenetic analysis. PCR products amplified with the archaeal-specific primer Arch21F (ref. 3) and the universal primer Univ1492R (ref. 24) were cloned using the TOPO TA Cloning Kit (Invitrogen). Sequences were obtained using ABI 3730XL (Applied Biosystems) or MegaBACE1000 (Amersham) automated DNA sequencers. Phylogenetic relationships were analysed by evolutionary distance and parsimony methods using the ARB software package²⁵. Evolutionary distances were calculated with the Kimura two-parameter distance correction. Regions of ambiguity were not included in

the analysis. Confidence estimates of branching order were determined in PAUP* 4.0 (ref. 26). Short sequences (<1,000 bp) were added to the final tree using the parsimony tool in ARB.

Fluorescence *in situ* hybridization. Fluorescently labelled polyribonucleotide probes targeting crenarchaeal 16S rRNA sequences were synthesized by *in vitro* transcription of nearly full-length crenarchaeal 16S ribosomal DNA clones amplified from enrichment cultures as previously described¹⁵. Whole-cell hybridizations and calculations of archaeal cell abundances were carried out on cells filtered onto 0.2-µm polycarbonate GTTP membranes (Millipore) following established protocols¹⁵. The specificity of the labelling was confirmed by the failure to hybridize against *Escherichia coli* and *Desulfovibrio vulgaris* cells. **Transmission electron microscopy.** Concentrated cells of SCM1 were fixed with 1% glutaraldehyde. The fixed sample suspension was diluted 1:1 with 2% uranyl acetate in water as droplets on Parafilm. A 200-mesh formvar-coated copper grid was floated on each cell/stain droplet for 5 min. Liquid was drained off with filter paper and the grid was air-dried. Each preparation was examined and photographed with a JEM 1200EXII transmission electron microscope (JEOL) at an accelerating voltage of 80 kV and magnifications of $\times 30,000$ and $\times 60,000$.

Scanning electron microscopy. Glutaraldehyde-fixed cell suspensions were applied to glass coverslips coated with poly-L-lysine, allowed to adhere for 5 min, rinsed with water and post-fixed with 1% OsO₄ for 30 min. Cells on coverslips were then rinsed with water, dehydrated through a graded ethanol series to absolute ethanol, and then critically point dried with L-CO₂. Coverslips were mounted on an aluminium stub, sputter-coated with Au/Pd and examined with a JSM 6300F scanning electron microscope (JEOL) at an accelerating voltage of 15 kV and magnifications of $\times 20,000$ –60,000.

Cloning of putative ammonia monoxygenase genes. Putative ammonia monoxygenase gene sequences from soil and planktonic mesophilic Crenarchaeota^{11,14} were aligned and used to design degenerate oligonucleotides for PCR amplification of *amoA* (CrenAmo1F, 5'-AATGGTCTGGCTWAGACGC-3'; CrenAmo1R, 5'-GACCARGCGGCCATCCA-3'), *amoB* (CrenAmo2.1F, 5'-CACGGTGTGTC CAAGCACA-3'; CrenAmo2.2R, 5'-RATTACYTGCCAVGGTC-3') and *amoC* (CrenAmo3.1F, 5'-ATGGCACARATGCCSGC-3'; CrenAmo3R, 5'-GGTATWG ATCTGCTGTACAA-3'). All PCR reactions were carried out as described for the 16S rRNA genes, except that 2.5 mM MgCl₂ was used in the PCR reaction mixture and the template DNA was 50 ng of purified genomic DNA from SCM1. All amplified AMO-encoding genes were cloned and sequenced as described above.

Received 21 March; accepted 8 June 2005.

- Burggraf, S., Huber, H. & Stetter, K. O. Reclassification of the crenarchaeal orders and families in accordance with 16S rRNA sequence data. *Int. J. Syst. Bacteriol.* **47**, 657–660 (1997).
- Fuhrman, J. A., McCallum, K. & Davis, A. A. Novel major archaeobacterial group from marine plankton. *Nature* **356**, 148–149 (1992).
- DeLong, E. F. Archaea in coastal marine environments. *Proc. Natl Acad. Sci. USA* **89**, 5685–5689 (1992).
- Dawson, S. C., Pace, N. R. & DeLong, E. F. Phylogenetic and ecological perspectives on uncultured Crenarchaeota and Korarchaeota. In *The Prokaryotes—An Evolving Electronic Resource for the Microbiological Community* 3rd edn (eds Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) (<http://link.springer-ny.com/link/service/books/10125/>) release 3.3 (Springer, New York, 2000).
- Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**, 507–510 (2001).
- Pearson, A., McNichol, A. P., Benitez-Nelson, B. C., Hayes, J. M. & Eglinton, T. I. Origins of lipid biomarkers in Santa Monica Basin surface sediment: A case study using compound-specific $\Delta^{14}\text{C}$ analysis. *Geochim. Cosmochim. Acta* **65**, 3123–3137 (2001).
- Wuchter, C., Schouten, S., Boschker, H. T. & Sinninghe Damste, J. S. Bicarbonate uptake by marine Crenarchaeota. *FEMS Microbiol. Lett.* **219**, 203–207 (2003).
- Ouverney, C. C. & Fuhrman, J. A. Marine planktonic Archaea take up amino acids. *Appl. Environ. Microbiol.* **66**, 4829–4833 (2000).
- Bock, E. & Wagner, M. Oxidation of inorganic nitrogen compounds as an energy source. In *The Prokaryotes—An Evolving Electronic Resource for the Microbiological Community* 3rd edn (eds Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E.) (<http://link.springer-ny.com/link/service/books/10125/>) release 3.7 (Springer, New York, 2001).
- Hovanec, T. A. & DeLong, E. F. Comparative analysis of nitrifying bacteria associated with freshwater and marine aquaria. *Appl. Environ. Microbiol.* **62**, 2888–2896 (1996).
- Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66–74 (2004).
- Preston, C. M., Wu, K. Y., Molinski, T. F. & DeLong, E. F. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc. Natl Acad. Sci. USA* **93**, 6241–6246 (1996).
- Béjà, O. *et al.* Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl. Environ. Microbiol.* **68**, 335–345 (2002).
- Treusch, A. H. *et al.* Characterization of large-insert DNA libraries from soil for environmental genomic studies of Archaea. *Environ. Microbiol.* **6**, 970–980 (2004).
- DeLong, E. F., Taylor, L. T., Marsh, T. L. & Preston, C. M. Visualization and enumeration of marine planktonic archaea and bacteria by using polyribonucleotide probes and fluorescent *in situ* hybridization. *Appl. Environ. Microbiol.* **65**, 5554–5563 (1999).
- Capone, D. G. in *Microbial Ecology of the Oceans* (ed. Kirchman, D. L.) 455–493 (Wiley, New York, 2000).
- Ducklow, H. in *Microbial Ecology of the Oceans* (ed. Kirchman, D. L.) 85–120 (Wiley, New York, 2000).
- Hügler, M., Huber, H., Stetter, K. O. & Fuchs, G. Autotrophic CO₂ fixation pathways in archaea (Crenarchaeota). *Arch. Microbiol.* **179**, 160–173 (2003).
- Menendez, C. *et al.* Presence of acetyl-coenzyme A (CoA) carboxylase and propionyl-CoA carboxylase in autotrophic Crenarchaeota and indication for operation of a 3-hydroxypropionate cycle in autotrophic carbon fixation. *J. Bacteriol.* **181**, 1088–1098 (1999).
- Wada, E. & Hattori, A. Nitrite metabolism in the euphotic layer of the central North Pacific Ocean. *Limnol. Oceanogr.* **16**, 766–772 (1971).
- Barns, S. M., Delwiche, C. F., Palmer, J. D. & Pace, N. R. Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl Acad. Sci. USA* **93**, 9188–9193 (1996).
- Pearson, A. *et al.* Nonmarine crenarchaeol in Nevada hot springs. *Appl. Environ. Microbiol.* **70**, 5229–5237 (2004).
- Widdel, F. & Bak, F. in *The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Application* 2nd edn (eds Ballows, A., Trüper, H. G., Dworkin, M., Harder, W. & Schleifer, K.-H.) 3352–3378 (Springer, New York, 1992).
- Lane, D. J. in *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M.) 115–175 (Wiley, New York, 1991).
- Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
- Swofford, D. L. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) (version 4.0 beta 10) (<http://paup.csit.fsu.edu/index.html>) (Sinauer, Sunderland, 2003).
- Stickland, J. D. H. & Parsons, T. R. *A Practical Handbook of Seawater Analysis* (Fisheries Research Board of Canada, Ottawa, 1972).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. R. Blevins, P. M. Berube and N. Pinel for technical assistance, S. Lara for performing electron microscopy and J. F. Heidelberg for assistance navigating the Sargasso Sea metagenome sequence data. We thank the Shedd and Seattle aquariums for samples and J. Hayes for his assistance. This research was supported by National Science Foundation Systematics (D.A.S.), Microbial Observatories (D.A.S. and J.B.W.) and Postdoctoral Fellowship (A.E.B.) programmes.

Author Information The sequences described in this manuscript have been deposited in GenBank under accession numbers DQ085097 to DQ085105. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.A.S. (dastahl@u.washington.edu).

Endangered plants persist under phosphorus limitation

Martin J. Wassen^{1*}, Harry Olde Venterink^{2*}, Elena D. Lapshina^{3†} & Franziska Tanneberger⁴

Nitrogen enrichment is widely thought to be responsible for the loss of plant species from temperate terrestrial ecosystems. This view is based on field surveys and controlled experiments showing that species richness correlates negatively with high productivity^{1,2} and nitrogen enrichment³. However, as the type of nutrient limitation has never been examined on a large geographical scale the causality of these relationships is uncertain. We investigated species richness in herbaceous terrestrial ecosystems, sampled along a transect through temperate Eurasia that represented a gradient of declining levels of atmospheric nitrogen deposition—from $\sim 50 \text{ kg ha}^{-1} \text{ yr}^{-1}$ in western Europe to natural background values of less than $5 \text{ kg ha}^{-1} \text{ yr}^{-1}$ in Siberia⁴. Here we show that many more endangered plant species persist under phosphorus-limited than under nitrogen-limited conditions, and we conclude that enhanced phosphorus is more likely to be the cause of species loss than nitrogen enrichment. Our results highlight the need for a better understanding of the mechanisms of phosphorus enrichment, and for a stronger focus on conservation management to reduce phosphorus availability.

Alterations to the environment by humans have reduced the plant species diversity in many ecosystems, and in some cases these reductions have affected ecosystem functioning⁵. To counteract this loss of diversity, there is an urgent need to uncover the underlying mechanisms responsible. Studies of diversity–productivity patterns in Canadian and European terrestrial wetlands suggest that increased productivity is a major factor influencing species extinction^{1,2}. According to Grime's 'hump-backed' model, there is a critical level of productivity at which species richness reaches a peak, and above which it declines rapidly because all but a few fast-growing, tall species are unsuccessful in competing for light⁶. Low or moderate nutrient availability has been thought to be a mechanism that reduces the competitive advantage of fast-growing, tall species relative to smaller ones⁷.

Aquatic freshwater ecosystems are generally thought to be P-limited⁸; although in freshwater lakes, species may also be lost due to nitrate enrichment⁹. In contrast, most terrestrial ecosystems of the temperate zone are considered to be N-limited^{10,11}; therefore, N-enrichment is seen as a major cause of plant species loss in temperate grasslands³ and forests¹¹. Although productivity only increases with increasing availability of the limiting resource¹², the type of nutrient limitation has never been examined for a large number of sites.

Here we provide evidence that P- rather than N-enrichment is more important in the loss of plant species from some ecosystems. We investigated the plant species composition of 274 sites with herbaceous vegetation, ranging from terrestrial freshwater wetlands

(fens, bogs and fluvial marshes) to moist grasslands. The sites—all in temperate Eurasia ($51\text{--}57^\circ \text{N}$)—were scattered along a west–east transect from the Netherlands/Belgium (5°E) through eastern Poland (23°E) to western Siberia (85°E). This transect also represents a gradient of declining levels of atmospheric nitrogen deposition, from high in the Netherlands and Belgium ($40\text{--}60 \text{ kg N ha}^{-1} \text{ yr}^{-1}$), to much lower in Poland ($5\text{--}10 \text{ kg N ha}^{-1} \text{ yr}^{-1}$), to very low in Siberia ($<5 \text{ kg N ha}^{-1} \text{ yr}^{-1}$; ref. 4). For all sites we recorded species richness of vascular plants, the number of endangered species (using the Dutch Red List¹³; see Methods) and the above-ground standing crop of vegetation. The type of nutrient limitation was inferred by analysing plant material and calculating nutrient ratio values^{14,15}.

As already indicated, if species loss is the result of increased productivity, then we would expect the endangered species in our samples to occur mainly at sites of low productivity. In addition, if the main cause of higher productivity is enhanced N deposition, then we can expect two further patterns to emerge. In western Europe we expect endangered species to be restricted to P-limited ecosystems. This is because N-enrichment will have transformed formerly unproductive, N-limited ecosystems into either highly productive ecosystems or low-productive P-limited ecosystems. In Poland and Siberia (where N-deposition is low) we expect endangered species to occur at sites of low productivity, irrespective of the type of nutrient limitation.

Two well established relationships are confirmed from our data. Firstly, the species richness–productivity relationships show the classical hump pattern^{1,2,6}, with highest species richness at productivities of $200\text{--}600 \text{ g m}^{-2}$ (Fig. 1a). Secondly, the sites with intermediate tissue N:P ratios (6–20) are on average the most species rich¹⁶ (Fig. 1b). Examining the patterns of endangered species shows that some aspects of our results seem to support the above mentioned expectations: (1) endangered plant species only occur at low-productivity sites (biomass $<\sim 600 \text{ g m}^{-2}$; Fig. 1c), (2) in the Netherlands/Belgium 70% of the endangered species occur at P-limited sites (Fig. 2a), and (3) in Poland and Siberia many endangered species are indifferent to the type of nutrient limitation (Fig. 2a). However, other findings are inconsistent with the initial hypotheses. For example, in Poland and Siberia endangered species are more frequent in P-limited sites than in N-limited sites (Fig. 2a), and in Poland this cannot be because P-limited sites are more frequent ($P < 0.05$, Fig. 2). Even more unexpectedly, most of the low-productivity sites in the Netherlands/Belgium are limited not by P but by N (Fig. 2b). Furthermore, maximum numbers of endangered species are higher under conditions of P- than N-limitation (indicated by the height of the red and blue 'peaks' in Fig. 1d, with the peak of the regression being at N:P

¹Environmental Sciences, Copernicus Institute for Sustainable Development and Innovation, Utrecht University, PO Box 80115, 3508 TC Utrecht, The Netherlands. ²Geobotanical Institute ETH, Zürichbergstrasse 38, 8044 Zürich, Switzerland. ³Department of Botany, State University of Tomsk, 36 Lenin Avenue, 634050, Tomsk, Russia. ⁴Institute of Botany and Landscape Ecology, Ernst-Moritz-Arndt University, Grimmer Strasse 88, 17487 Greifswald, Germany. [†]Present address: Yugra State University, Department of Environmental Dynamics and Global Climate Change, ul. Chekhova 16, 628012, Khanty-Mansiysk, Russia.

*These authors contributed equally to this work.

ratio 21). Moreover, the percentage of endangered species in the vegetation clearly increases with increasing P-limitation (Fig. 1f). Hence, endangered plant species appear better able to persist in P-limited sites than in sites with other types of nutrient limitation.

Three different mechanisms may explain the observed pattern: (1) there is a wider variation in adaptations to low P-availability (for example, phosphatase exudation, soil acidification, cluster roots or mycorrhiza) than to low N-availability (for example, symbiotic N_2 -fixation or organic N-uptake). This may have resulted in a larger pool of species adapted to low P-availability; (2) low productive systems that used to be N-limited have become P-limited because of N-enrichment; or (3) human impact has impaired P-limited ecosystems more strongly than N-limited ecosystems, leading to a larger loss of species adapted to low P-availability.

The first explanation is unlikely because if species pools differed between N-limited and P-limited sites, then the total species richness would be expected to differ, which is not the case (Fig. 1a, b). The second explanation must also be discarded because in the Netherlands and Belgium (where N deposition is high) N-limited ecosystems with low productivity are more common than P-limited ecosystems (Fig. 2b). The third explanation seems the most plausible, because in western Europe various human impacts have enhanced P-availability in wet and moist ecosystems. Increased groundwater

extraction for drinking water and industry have diminished the discharge of calcium- and iron-rich groundwater into wetlands and consequently this has reduced the binding of $P^{17,18}$. Together with this effect, P-enrichment of surface waters¹⁹ and internal eutrophication (through for instance increased drying–wetting dynamics²⁰ or sulphate pollution²¹) of the soil have also contributed to enhanced P-availability. On a global scale, it has been estimated that human intervention in the biogeochemical P cycle has increased the magnitude of P fluxes by 400%, which is far more than for carbon, nitrogen and sulphur²². We conclude that, in spite of severely enhanced atmospheric N-deposition in Western Europe⁴, P-enrichment has been more important than N-enrichment for species loss from wet and moist herbaceous ecosystems. P-enrichment may have caused productivity increase and species loss through competitive exclusion, or a shift from P- to N-limitation, to the disadvantage of species adapted to low P availability. In a discussion of whether species richness can ever be raised by nutrient addition, Güsewell and colleagues¹⁶ suggested that addition of P may increase the total

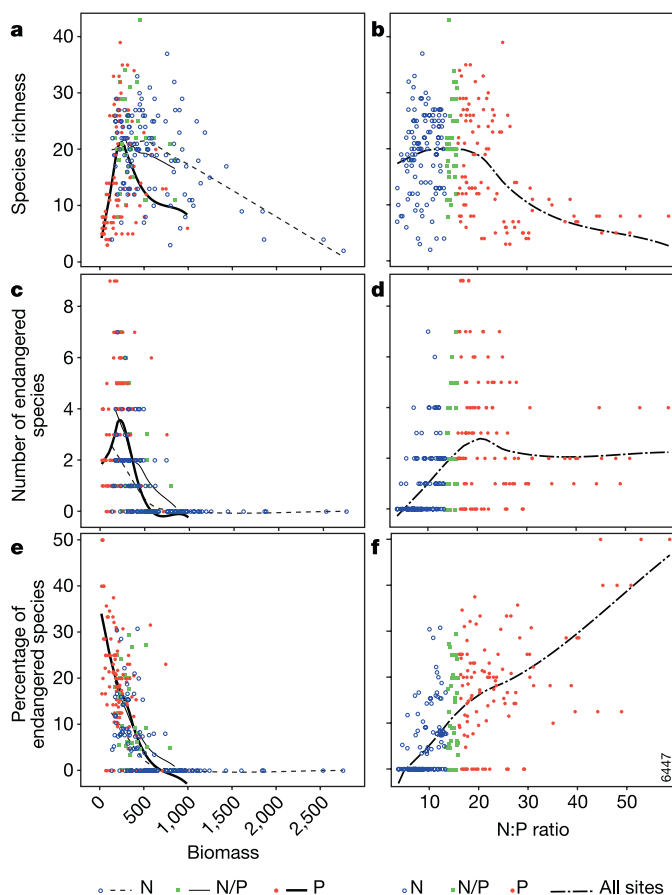


Figure 1 | Species richness affected by nutrient limitation in herbaceous ecosystems in Eurasia. a–f, The total number of vascular plant species (a and b), endangered species among them (c and d) and endangered species as percentage of all vascular species (e and f) are plotted against above-ground biomass (in $g\ m^{-2}$) of vascular plants (a, c, e), and N:P ratio in above-ground vascular plant material (b, d, f). Blue circles are N-limited sites, red filled circles are P-limited sites and green filled squares are N/P co-limited sites. Trends were analysed using LOWESS regression (span 2/3; degree 1). For f we also performed a linear regression ($y = 0.77x - 0.94$; $R^2 = 0.40$; $P < 0.001$).

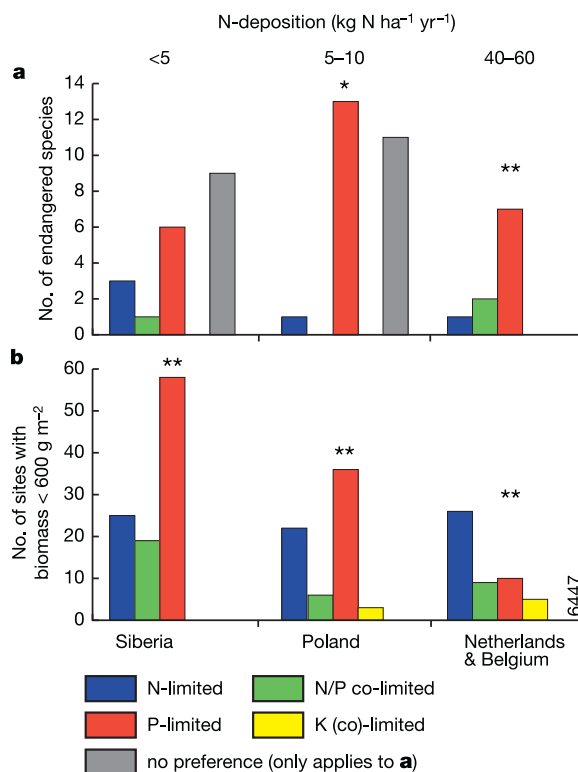


Figure 2 | Frequency distributions of endangered species and sites with low productivity. a, b, Frequency distributions of endangered species with preference for a certain type of nutrient limitation per region (a), as well as frequency distribution of sites with biomass $< 600\ g\ m^{-2}$, that is, the biomass range where endangered species occur (b). Preference of species α for example, P-limitation means that species α occurred only at P-limited sites or it occurred at least three times more often at P-limited sites than at sites limited by another nutrient (N, N/P or K). Species not showing preference according to this definition were labelled as no preference. In b, the observed distributions all differ significantly from an equal frequency distribution among the various types of nutrient limitation (χ^2 test; two asterisks indicates $P < 0.001$). In a, we tested whether the observed distribution of endangered species among N-, N/P-, P- and K-limited sites differs significantly from the frequency patterns shown in b (that is, distributions that could be expected on the basis of frequency of occurrence of N-, N/P-, P- and K-limited sites assuming endangered species had equal probabilities for occurring on N-, N/P-, P- and K-limited sites ($P = 0.25$)). For the Polish and Dutch/Belgian sites, the observed distribution differs significantly from expected distributions, for the Siberian sites it does not (χ^2 test; one asterisk indicates $P < 0.05$; two asterisks indicate $P < 0.001$).

species richness of P-limited wetlands, but the authors warned that this would tend to promote common species at the expense of rare ones. Our result fully supports their expectation; regardless of patterns in total species richness, both the absolute number and the proportion of endangered species in the vegetation appeared to be greatest in P-limited wetlands.

Our findings suggest that P-limitation in terrestrial systems may be more widespread than generally acknowledged²³, and imply that the conservation of endangered species requires the preservation and restoration of P-limited ecosystems. Conservation managers often attempt to enhance the plant species diversity of an ecosystem by reducing its nutrient capital, and thereby its productivity. Although these attempts are sometimes successful²⁴, productivity reduction often does not lead to an increase in species richness^{25,26} because conditions of P-limitation are not restored. For example, re-establishment of endangered species on former agricultural fields generally fails²⁷ because of the large P-pool accumulated in the soil over decades of fertilization²⁸. In general, policies biased towards reducing nitrogen enrichment (for example, the European Union Directive on Nitrate) are unlikely to provide adequate protection for the majority of endangered species in herbaceous ecosystems. A multivariate consideration of anthropogenic impacts on the water cycle, the biogeochemical cycles (of C, N, P and S) and soil acidity is needed. This requires a systems approach²² that includes the assessment of the effect of changes in ecological stoichiometry on biodiversity.

METHODS

Sampling protocol. We sampled 150 plots as described in ref. 15. In 2001, 2002 and 2003 we sampled 124 additional plots of 10 m² following the same method. The data set contains terrestrial wetlands (for example, river marginal floodplains and marshes, wet grasslands, fens, bogs and fen meadows) as well as moist grasslands. Within each plot we randomly sampled a subplot of 0.16 m² of which we harvested above-ground standing crop at the height of the growing season (July). We separated mosses and vascular plants and determined dry weight and contents of N, P and K in plant material (Kjeldahl destruction). We used dry weight of above-ground biomass of vascular plants as an estimate of primary production (see refs 15 and 29).

Determining the type of nutrient limitation. To determine the type of nutrient limitation we used a method based on critical values of N:P, N:K and K:P ratios in above-ground plant material derived from literature reviews of fertilization experiments^{14,15}. N- and P-limited sites were distinguished on the basis of N:P ratios with N:P ratios >16 indicating P-limitation, N:P ratios <13.5 indicating N-limitation, and between 13.5 and 16 indicating N/P co-limitation¹⁴. This N/P co-limitation should be interpreted as true co-limitation by N and P together, or at least as no clear single limitation by N or P. For distinguishing K (co)-limited samples we used the critical N:K ratio of 2.1 and the critical K:P ratio of 3.4 (ref. 15). Only 11 sites were K (co)-limited (see Supplementary Table 1). These sites are excluded from Fig. 1.

The data set presented here includes different types of herbaceous vegetation, some herb-rich and others dominated by grasses. As grasses might have intrinsically higher N:P ratios than herbs¹⁴, differences in dominance of these plant groups could have affected average N:P ratios of a site. To evaluate whether dominance by either grasses or herbs has an influence on N:P ratios and the occurrence of endangered species we divided the data into three parts: (1) sites in which grasses accounted for more than half of the plant cover; (2) sites in which herbs accounted for more than half of the plant cover, and (3) sites in which neither grasses nor herbs accounted for more than half of the plant cover (excluding the 11 K-limited sites; $n = 263$). We tested whether N:P ratios differed significantly among these three groups (Tukey test after one-way analysis of variance, ANOVA). The majority of sites was neither dominated by grasses nor by herbs and only 9% of the sites in the biomass range where endangered species occur (<600 g m⁻²) were dominated by grasses (see Supplementary Table 2). The grass:herb ratio does affect N:P ratio, with grass-dominated sites having a significantly lower N:P ratio. However, in the subset where endangered species occur the N:P ratio of the sites dominated by grasses did not differ significantly from that of the sites dominated by herbs. Additional information and discussion concerning the method for assessing nutrient limitation by means of N:P ratios is given in the Supplementary Discussion.

Plant species. We analysed the similarity between the list of species recorded by us in Poland and Russia and the Dutch flora. Eighty seven percent of the species of the Polish sites, and 75% of the species of the west Siberian sites belong to the

Dutch flora. From the Red List of endangered species of the Netherlands¹³, we only used 'actually threatened species', that is, species that have disappeared in at least 25% of the map units (1 unit is 25 km²) between 1970 and 1990. The majority of the endangered species of the Netherlands also occur on red lists of other western and central European countries; that is, 100%, 80%, 71%, 67%, 63% and 27% of the species observed by us which are on the Dutch Red List (see Supplementary Table 1) also occur on the Red Lists of Flanders, Germany, Czech Republic, Switzerland, Luxembourg and Poland, respectively.

Other factors affecting diversity-productivity patterns. Species diversity in the data set is affected by stress factors such as acidity and inundation which operate through sizes of regional species pools³⁰ (see Supplementary Fig. 2). However, these factors cannot explain our main result, more endangered species at P-limited sites. The type of nutrient limitation was not related to acidity, and apart from some sites (riparian marshes deeply flooded in spring) nor to the maximum inundation depth. Other stress factors like salinity and drought stress did not play a role in our data set.

Received 20 May; accepted 23 June 2005.

- Moore, D. R. J., Keddy, P. A., Gaudet, C. L. & Wisheu, I. C. Conservation of wetlands: do infertile wetlands deserve a higher priority? *Biol. Conserv.* **47**, 203–217 (1989).
- Wheeler, B. D. & Shaw, S. C. Above-ground crop mass and species richness of the principal types of herbaceous rich-fen vegetation of lowland England and Wales. *J. Ecol.* **79**, 285–301 (1991).
- Stevens, C. J., Dise, N. B., Mountford, J. O. & Gowing, D. J. Impact of nitrogen deposition on the species richness of grasslands. *Science* **303**, 1876–1879 (2004).
- Holland, E. A., Dentener, F. J., Braswell, B. H. & Sulzman, J. M. Contemporary and pre-industrial global reactive nitrogen budgets. *Biogeochemistry* **46**, 7–43 (1999).
- Naeem, S., Thompson, J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
- Grime, J. P. *Plant Strategies and Vegetation Processes* (Wiley, Chichester, 1979).
- Taylor, D. R., Aarssen, L. W. & Loehle, C. On the relationship between r/K selection and environmental carrying capacity: a new habitat template for plant life history strategies. *Oikos* **58**, 239–250 (1990).
- Schlesinger, W. H. *Biogeochemistry: An Analysis of Global Change* (Academic, San Diego, 1997).
- Gilles, J. G. Nitrogen study fertilizes fears of pollution. *Nature* **433**, 791 (2005).
- Vitousek, P. M. & Howarth, R. W. Nitrogen limitation on land and in the sea: how can it occur? *Biogeochemistry* **13**, 87–115 (1991).
- Sala, O. E. et al. Global biodiversity scenarios for the year 2100. *Science* **287**, 1770–1774 (2000).
- Chapin, F. S. I., Vitousek, P. M. & van Cleve, K. The nature of nutrient limitation in plant communities. *Am. Nat.* **127**, 48–58 (1986).
- van der Meijden, R., van Duuren, L., Weeda, E. J. & Plate, C. L. Standaardlijst van de Nederlandse flora 1990. *Gorteria* **17**, 75–127 (1991).
- Güsewell, S. & Koerselman, W. Variation in nitrogen and phosphorus concentrations of wetland plants. *Perspect. Plant Ecol. Evol. Syst.* **5**, 37–61 (2002).
- Olde Venterink, H., Wassen, M. J., Verkoort, A. W. M. & de Ruiter, P. C. Species richness-productivity patterns differ between N-, P- and K-limited wetlands. *Ecology* **84**, 2191–2199 (2003).
- Güsewell, S., Bailey, K. B., Roem, W. J. & Bedford, B. L. Nutrient limitation and botanical diversity in wetlands: can fertilization raise species richness? *Oikos* **109**, 71–80 (2005).
- Wassen, M. J., van Diggelen, R., Wolejko, L. & Verhoeven, J. T. A. A comparison of fens in natural and artificial landscapes. *Vegetatio* **126**, 5–26 (1996).
- Richardson, C. J. Mechanisms controlling phosphorus retention capacity in freshwater wetlands. *Science* **228**, 1424–1427 (1985).
- Carpenter, S. R. et al. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecol. Appl.* **8**, 559–568 (1998).
- Turner, B. L. & Haygarth, P. M. Phosphorus solubilization in rewetted soils. *Nature* **411**, 258 (2001).
- Lamers, L. P. M., Tomassen, H. B. M. & Roelofs, J. G. M. Sulphate induced eutrophication and phytotoxicity in freshwater wetland. *Environ. Sci. Technol.* **32**, 199–205 (1998).
- Falkowski, P. et al. The global carbon cycle: a test of our knowledge of Earth as a system. *Science* **290**, 291–296 (2000).
- Elser, J. J. et al. Nutritional constraints in terrestrial and freshwater food webs. *Nature* **408**, 578–580 (2000).
- Bakker, J. P. *Nature Management by Grazing and Cutting* (Kluwer Academic, Dordrecht, 1989).
- Berendse, F., Oomes, M. J. M., Altena, H. J. & Elberse, W. T. Experiments on the restoration of species-rich meadows in the Netherlands. *Biol. Conserv.* **62**, 59–65 (1992).
- Jansen, A. J. M. & Roelofs, J. G. M. Restoration of *Cirsio-Molinietum* wet meadows by sod cutting. *Ecol. Eng.* **7**, 279–298 (1996).

27. Bakker, J. P. & Berendse, F. Constraints in the restoration of ecological diversity in grassland and heathland communities. *Trends Ecol. Evol.* **14**, 63–68 (1999).
28. Gough, M. W. & Marrs, R. H. A comparison of soil fertility between semi-natural and agricultural plant communities: implications for the creation of species-rich grassland on abandoned agricultural land. *Biol. Conserv.* **51**, 83–96 (1990).
29. Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127 (1999).
30. Grace, J. B. The factors controlling species density in herbaceous plant communities: an assessment. *Perspect. Plant Ecol. Evol. Syst.* **2**, 1–28 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank A. Barendregt, D. Belgers, W. Bleuten, W. Hahne, A. de Hamer, H. de Mars, N. Pieterse, S. Schep, J. de Smidt, R. Zeefat and J. van Zuidam for their assistance during the fieldwork, P. de Ruiter, M. Rietkerk, S. Güsewell, H. Joosten, P. Edwards and N. Versfeld for their comments on the manuscript, V. Rodenburg for reading the manuscript, W. Busschers for statistical advice and T. Markus for drawing the figures.

Author Contributions M.J.W. project planning; M.J.W. and H.O.V. data analysis and writing; M.J.W., H.O.V., E.D.L. and F.T. field work.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.W. (m.wassen@geo.uu.nl).

DNA sequence and analysis of human chromosome 18

Chad Nusbaum¹, Michael C. Zody¹, Mark L. Borowsky¹, Michael Kamal¹, Chinnappa D. Kodira¹, Todd D. Taylor², Charles A. Whittaker^{1†}, Jean L. Chang¹, Christina A. Cuomo¹, Ken Dewar^{1†}, Michael G. Fitzgerald¹, Xiaoping Yang¹, Amr Abouelleil¹, Nicole R. Allen¹, Scott Anderson¹, Toby Bloom¹, Boris Bugalter¹, Jonathan Butler¹, April Cook¹, David DeCaprio¹, Reinhard Engels¹, Manuel Garber¹, Andreas Gnirke¹, Nabil Hafez¹, Jennifer L. Hall¹, Catherine Hosage Norman¹, Takehiko Itoh³, David B. Jaffe¹, Yoko Kuroki², Jessica Lehoczy^{1†}, Annie Lui¹, Pendexter Macdonald¹, Evan Mauceli¹, Tarjei S. Mikkelsen¹, Jerome W. Naylor¹, Robert Nicol¹, Cindy Nguyen¹, Hideki Noguchi^{2,4}, Sinéad B. O'Leary¹, Bruno Piqani¹, Cherylyn L Smith¹, Jessica A. Talamas¹, Kerri Topham¹, Yasushi Totoki², Atsushi Toyoda², Hester M. Wain⁵, Sarah K. Young¹, Qiandong Zeng¹, Andrew R. Zimmer¹, Asao Fujiyama^{2,6}, Masahira Hattori^{2,7}, Bruce W. Birren¹, Yoshiyuki Sakaki² & Eric S. Lander¹

Chromosome 18 appears to have the lowest gene density of any human chromosome and is one of only three chromosomes for which trisomic individuals survive to term¹. There are also a number of genetic disorders stemming from chromosome 18 trisomy and aneuploidy. Here we report the finished sequence and gene annotation of human chromosome 18, which will allow a better understanding of the normal and disease biology of this chromosome. Despite the low density of protein-coding genes on chromosome 18, we find that the proportion of non-protein-coding sequences evolutionarily conserved among mammals is close to the genome-wide average. Extending this analysis to the entire human genome, we find that the density of conserved non-protein-coding sequences is largely uncorrelated with gene density. This has important implications for the nature and roles of non-protein-coding sequence elements.

The International Human Genome Sequencing Consortium (IHGSC) recently completed a sequence of the human genome and published a report on the finishing of the human genome^{2,3}. Now, papers containing detailed reports about each human chromosome are bringing to light aspects of the biomedical and evolutionary implications of this work. Here we describe the completion of a physical map, high-quality finished sequence, and gene catalogue for human chromosome 18, which represents approximately 2.7% of the human genome.

The extremely low density of protein-coding genes on chromosome 18 (Table 1) offers an opportunity to study the conservation of non-protein-coding sequences. It was recently observed that, in addition to protein-coding sequences, ~3% of the human genome shows a degree of evolutionary conservation among mammals that is significantly higher than background⁴. It is unclear whether this sequence consists mostly of regulatory elements related to genes or whether it represents other elements not tightly coupled to genes. These alternatives can be explored by comparing gene-rich and gene-poor chromosomes to see whether the proportion of conserved

non-protein-coding sequence tends to scale with gene density or is unrelated to gene density.

The finished sequence of chromosome 18 contains 76,117,153 bases and is interrupted by three euchromatic gaps, one gap at the 18q telomere and one gap containing the centromeric heterochromatin (Fig. 1 and Supplementary Table S2). These gaps are refractory to current cloning and mapping technology. The sizes of the euchromatic gaps were estimated by alignment to the regions of conserved synteny in the mouse genome⁴ (see Methods). The size of the telomeric gap was estimated using the size of the telomeric half-YAC (yeast artificial chromosome). The total size of these gaps is estimated at 118 kb. This corresponds to <0.2% of the euchromatic length of the chromosome, substantially lower than the average across the human genome (cited in ref. 3, also refs 5–7). Of the finished sequence, 79% was generated by the Broad Institute of MIT and Harvard (formerly the Whitehead Institute/MIT Center for Genome Research or WICGR), 20% by the RIKEN Genomic Sciences Center, and the remaining 1% by three other research groups (Supplementary Tables S3, S4). Details of construction of the clone map and sequencing are described in the Supplementary Information.

Several analyses verify that nearly the entire euchromatic region of chromosome 18 is present and accurately represented in the finished sequence. Of the 332 gene sequences in the well-curated RefSeq⁸ data set that have been mapped to chromosome 18, all are present and complete in the finished sequence. In addition, the finished sequence shows excellent alignment to genetic and radiation hybrid maps (Supplementary Fig. S1). The genetic map⁹ shows perfect alignment, with no discrepancies among 156 sequence-based genetic markers (Supplementary Table S5). The radiation hybrid map¹⁰ shows good agreement, but contains local discrepancies that would be expected from its lower resolution (Supplementary Table S6).

We assessed the local accuracy of the clone path by aligning paired-end sequences from a human Fosmid library (designated WIBR2,

¹Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, Massachusetts 02141, USA. ²RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan. ³Mitsubishi Research Institute Inc., 2-3-6 Otemachi, Chiyoda-ku, Tokyo 100-8141, Japan. ⁴University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa 277-0882, Japan. ⁵HUGO Gene Nomenclature Committee, The Galton Laboratory, Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK. ⁶National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. ⁷Kitasato Institute for Life Sciences, Kitasato University 1-15-1, Kitasato, Sagami-hara, Kanagawa 228-8555, Japan. [†]Present addresses: MIT Center for Cancer Research, 77 Mass Avenue E18-570, Cambridge, Massachusetts 02139, USA (C.A.W.). McGill University and Genome Quebec Innovation Centre, Montreal, Quebec H3A 1A4, Canada (K.D.). Department of Human Genetics, University of Michigan Medical School, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109, USA (J.L.).

Table 1 | Chromosome 18 gene content

	Gene no.	Gene %	Gene length (bp) *	No. of alternative transcripts	Transcript length (bp)†	No. of exons per transcript‡	Internal exon length (bp)§	Intron length (bp)	CpG-5' association¶
Known genes	243	72	88,523	3.1	3,121	10.7	155 (<i>n</i> = 2,351)	9,068 (<i>n</i> = 2,997)	73
Novel transcripts	10	6	44,778	1.2	870	4.6	146 (<i>n</i> = 68)	7,131 (<i>n</i> = 101)	33
Putative genes	11	3	10,427	1.0	560	2.2	97 (<i>n</i> = 4)	6,425 (<i>n</i> = 18)	36
Novel CDS	49	15	58,294	1.9	987	4.4	145 (<i>n</i> = 217)	12,150 (<i>n</i> = 288)	0
Gene fragments	13	4	2,095	1.0	2,027	1.0			0
PredictedPlus	11	3	57,060	1.0	1,008	5.6	137 (<i>n</i> = 40)	12,576 (<i>n</i> = 49)	18
Total	337		75,519						
Pseudogenes	171	51	3,601	1.0	837	1.9	178 (<i>n</i> = 105)	2,971 (<i>n</i> = 159)	7

* Average chromosomal distance from start of 5'-most exon to 3'-most exon for all transcripts of a gene.

† Average length summed across the footprint of all exons for all transcripts of a gene—total exon space per gene.

‡ Average number of exons in transcripts. Exons common to different transcripts were counted once per transcript.

§ Average length of exons using the footprint of all non-terminal exons for all transcripts of a gene. Unique overlapping exons or contained exons are counted separately, making this an average length of unique exons in a gene.

|| Average length of unique introns in a gene. In the case of exon skipping, both the shorter and longer overlapping introns were counted towards the average.

¶ Percentage of genes with a transcript having a CpG island (as assessed by FirstEF) within -2 kb and +1 kb of transcription start.

representing 10× physical coverage) to the finished sequence³. By identifying discrepancies in the distances between Fosmid ends in the finished sequence and those expected on the basis of insert size constraints, one can detect errors in the clone path³. Our analysis revealed a single aberrant region, which was found to result from a bacterial artificial chromosome (BAC) clone containing a 21-kb deletion that was either present in the source genome or occurred in the cloning of the BAC; this clone was replaced with a non-deleted BAC from a different library. Finally, an independent quality assessment exercise commissioned by NHGRI estimated the accuracy of the finished sequence at less than one error per 100,000 bases¹¹ (J. Schmutz, personal communication).

We produced a manually curated catalogue of genes (see Methods), annotating 337 gene loci and 171 pseudogene loci on chromosome 18. These include all previously known genes on chromosome 18 (Table 1). According to the Hawk2 categorization scheme (<http://www.sanger.ac.uk/Info/workshops/hawk2>, see Supplementary Information) there are 243 'known' genes, 49 'novel CDS' (coding sequence of a gene), 10 'novel transcripts', 11 'putative

genes', 11 'predictedplus genes' and 13 'gene fragments'. All 'novel transcript' genes had expressed-sequence-tag (EST) evidence. For 'putative genes', only a subset of the exons were supported by one or more spliced ESTs. Only a small fraction of all loci, those in the 'novel' and 'putative' categories, were annotated as genes on the basis of spliced EST evidence only. Some 'gene fragment' loci may prove to be pseudogenes.

Using aligned EST evidence, it was possible to extend many of the previously known gene models at their 5' or 3' ends (see Supplementary Fig. S2 for an example). Approximately 57% of the RefSeq and mammalian gene collection (MGC) transcripts could be extended. The 5' end extensions averaged 321 bp, and 3' end extensions averaged 1,131 bp. In addition, a novel 5' exon was found for 14% of the RefSeq or MGC transcripts, and a novel 3' exon was found for 2.2%. The ability to extend the gene models probably reflects expanded databases of transcripts and ESTs. A sampling of the extended gene models was validated in the laboratory (see Supplementary Information).

We found an average of 10.7 exons per full-length known

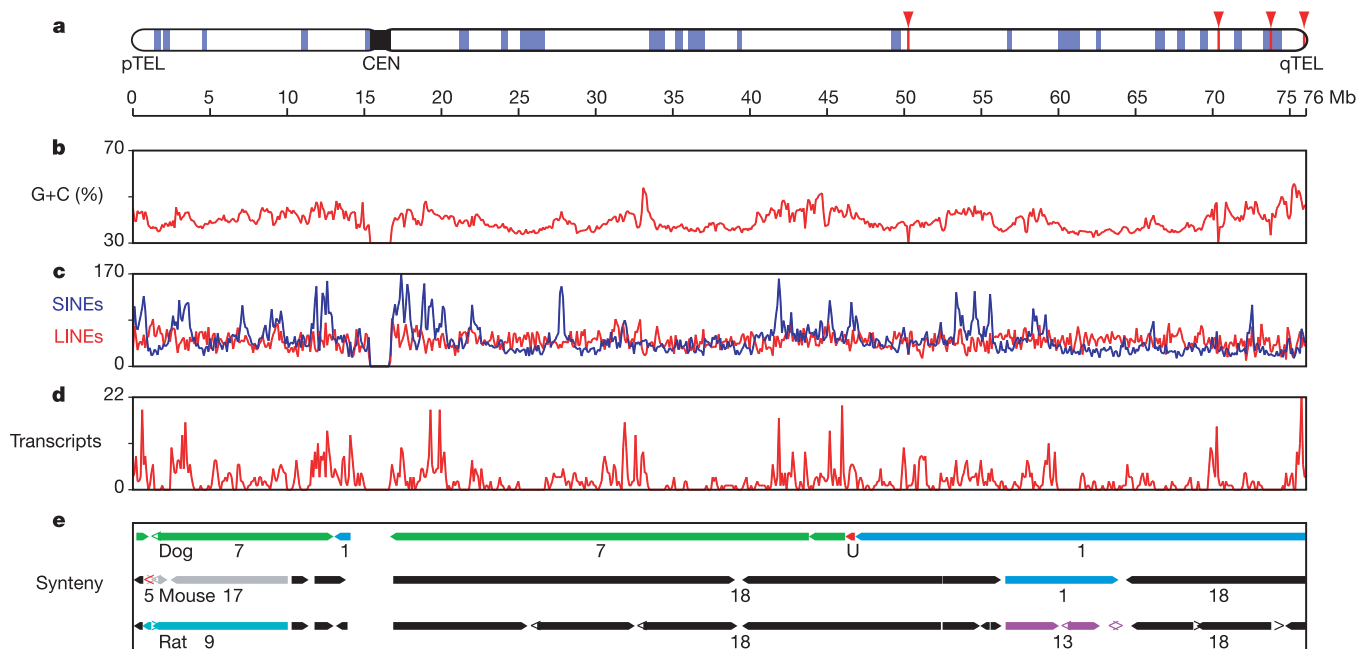


Figure 1 | Overview of human chromosome 18. **a**, Blue shading indicates gene deserts (≥ 500 kb with no transcript, see Supplementary Table S8). Telomeres (pTEL and qTEL), the centromere (CEN) and euchromatic sequence gaps (red lines) are also indicated. **b**, G+C content in discrete windows of 100 kb. **c**, **d**, Densities of long interspersed nuclear elements

(LINES, red), short interspersed nuclear elements (SINES, blue) and transcripts (**d**) are shown as numbers of these elements in discrete windows of 100 kb. **e**, Blocks of conserved synteny (100-kb resolution) with dog, mouse and rat, determined for this work. Chromosomes are numbered, and are coloured arbitrarily for ease of distinction.

transcript, comparable to recent published reports of human chromosomes. Internal exon lengths average 155 bp, and the average transcript length is 3.1 kb for full-length transcripts of known genes. There is evidence of extensive alternative splicing, with gene loci having an average of 3.1 distinct transcripts and 71% having at least two transcripts. This rate of alternative splicing is comparable to recent reports^{5,6}.

The longest gene on chromosome 18 is *DCC* (deleted in colorectal carcinoma), spanning 1,190,632 bp. *DCC* also contains the longest intron at 411,177 bp. The longest mature transcript is laminin $\alpha 3$ (*LAMA3*) at 10,585 bp. The longest single exon is found in *TCF4*, being a 3' exon of 5,700 bp. The gene with the most identified splice forms is *TGIF* (TGF β -induced factor), which appears to have ten splice forms, of which two are represented by RefSeq transcripts. Of the 171 pseudogenes on chromosome 18, approximately two-thirds are processed (intronless) pseudogenes arising from retroposition, and the remaining one-third are unprocessed. In addition, we identified four transfer RNA genes on the chromosome, listed in Supplementary Table S7. An analysis of gene families revealed that several families have multiple members present on chromosome 18. These include members of the laminin and cadherin families of cell adhesion molecules, and a cluster of ten serpin serine protease inhibitors (see Supplementary Information). Careful analysis of gene models found 59 pairs of overlapping genes on chromosome 18, suggesting that overlapping genes may be 2–4 times more common than previously thought^{12,13} (see Supplementary Information).

With an average of 4.4 genes per megabase (Mb), chromosome 18 has the lowest gene density of published human chromosomes (Supplementary Table S1). This gene density cannot be explained by chance fluctuation around a genome-wide mean ($P < 10^{-12}$, see

Supplementary Information). The low gene density is reflected both in the low percentage of transcribed sequence (28.5%) and the small fraction of the chromosome included in exons (1.14% in all exons, 1.06% in coding exons). The G+C content (39.8%) is also low, consistent with the known positive correlation between G+C content and gene number¹⁴.

Chromosome 18 contains 24 gene deserts (defined as a 500-kb region without a coding gene, Supplementary Table S8), which together comprise 28 Mb or $\sim 38\%$ of the total chromosome length. The sparsest region of the chromosome harbours only three genes over 4.5 Mb. In addition, chromosome 18 also has the longest median length of introns among all chromosomes, reflecting a genome-wide inverse correlation between intron size and gene density (Supplementary Fig. S3).

Despite being gene-poor, chromosome 18 is not enriched in repeat sequences. Transposable element fossils cover 43.5% of the chromosome, which is typical across the genome. Chromosome 18 also has a relatively low proportion of segmental duplication (segmental duplications are defined as having greater than 90% identity and being longer than 1 kb). Segmental duplications constitute $\sim 2.5\%$ (1.92 Mb) of the chromosome, with a greater representation of interchromosomal duplications (2.13%) than intrachromosomal duplications (0.55%). Some sequences are represented in both types of duplication (E. Eichler and X. She, personal communication).

The paucity of genes on chromosome 18 probably explains why it is one of only three autosomes (the others being chromosomes 13 and 21) for which trisomic individuals routinely survive to term¹ (www.trisomy.org, www.ndss.org). Although chromosomes 18 and 21 have roughly the same number of RefSeq genes (332 and 374 genes, respectively), chromosome 18 trisomy (Edwards syndrome) has much more severe health effects than chromosome 21 trisomy

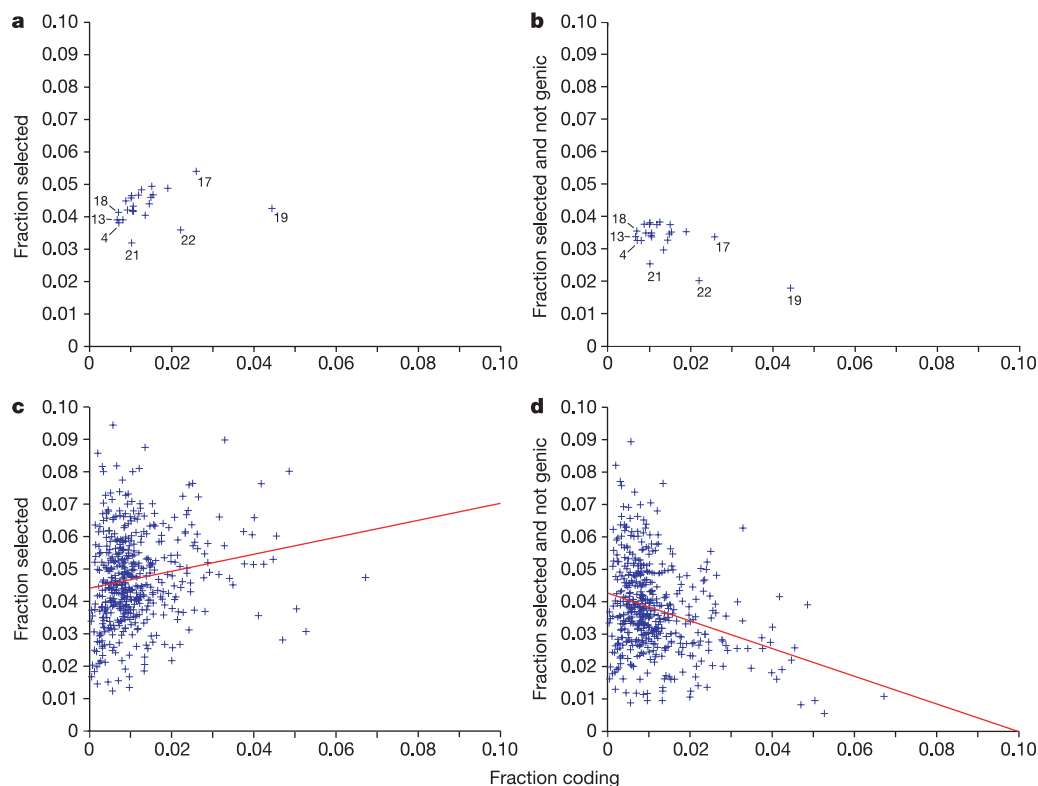


Figure 2 | Scatter plots showing the fraction of syntenic region under selection plotted against the fraction of coding sequence in that region. a, By chromosome, the fraction of all sequence under selection versus the coding fraction. b, By chromosome, the fraction of all non-protein-coding sequence under selection versus the coding fraction. Numbers refer to specific chromosomes. c, The fraction of all sequence under selection within

the region versus the coding fraction within the region. d, The fraction of all non-protein-coding sequence under selection versus the coding fraction. In c and d, each point represents a 5-Mb region from a set of non-overlapping 5-Mb regions covering the genome. Lines of regression are shown. We define non-protein-coding sequence as that which is completely disjoint from any predicted mature mRNA product of an annotated protein-coding gene.

(Down syndrome). Edwards syndrome occurs in 1 in 5,000 live births, and ~90% of affected individuals die before one year of age. In contrast, Down syndrome is more common (1 in 800 live births), and affected individuals are frequently able to cope with the numerous health consequences and survive to adulthood. The availability of gene catalogues for these two chromosomes will facilitate work to elucidate how the contributions of specific genes lead to such different clinical outcomes.

Four other syndromes are caused by gross abnormalities in chromosome 18, including three partial monosomies caused by deletion of part of the p or q arms (18p-, 18q- and ring18) and tetrasomy of the p arm (www.chromosome18.org). The gene catalogue presented here should facilitate identification of the critical genes associated with each syndrome.

At least 45 loci on chromosome 18 have been implicated in genetic disorders¹⁵ (Supplementary Table S9). The list includes at least four disorders for which the responsible gene and molecular mechanism of disease have been characterized (Supplementary Table S9). For two such diseases (methemoglobinemia and erythropoietic protoporphyria), we found evidence for novel alternative splice forms that would result in coding sequence alterations (not shown).

Comparative gene analysis revealed one locus that may represent a newly evolved gene in the primate lineage, although its function is unknown. Among the annotated multi-exon genes contained in blocks of conserved synteny among mammals, only one lacks exonic conservation with rodents and dog: *C18orf2*, a predicted RefSeq gene. Within this block of conserved synteny there is a primate-specific ~100-kb inversion in the region (present in both human and chimpanzee). One of the endpoints of this inversion lies in the middle of the coding region of the gene, with the result that the region is not contiguous in either dog or rodent genomes. Partial sequencing of this gene in apes suggests that it is conserved at least as far back as orangutan (see Supplementary Information).

We compared chromosome 18 to its homologue chimpanzee chromosome 18 (ref. 16). The average sequence divergence is 1.25%, which is close to the genome-wide average. On a larger scale, the karyotype of human chromosome 18 differs from its homologues in the great apes by a human-specific pericentric inversion with an associated human-specific inverted duplication of 19 kb (refs 17, 18). As a consequence, human 18p corresponds to the proximal region of chimpanzee 18q. As large-scale chromosomal rearrangements can facilitate speciation^{19,20}, it is possible that this inversion had had a role in hominid evolution.

Finally, we sought to explore the still-mysterious nature of conserved non-protein-coding sequences. Recent comparison of the human and mouse genomes⁴ led to the surprising discovery that ~5% of the human genome shows evolutionary conservation higher than the background rate (defined as the rate seen in ancestral repeat elements, which are presumed to be non-functional). Similar results have been seen in comparisons between the human and rat genomes²¹. As only 1–2% of the human genome encodes protein-coding exons, this indicates that the majority of human sequence under purifying selection is non-protein-coding. In principle, these non-protein-coding sequences could be (1) associated with protein-coding genes, such as those that directly or indirectly regulate the expression of protein-coding genes, or (2) independent of protein-coding genes, such as those that play a structural role in chromosome architecture or those that encode RNA genes.

We calculated the overall proportion of bases on each chromosome that are under purifying selection, and allocated this proportion as either protein-coding or non-protein-coding (see Methods). The computational analysis closely followed that used in recent mammalian comparisons^{4,22} (see Methods). We compared the proportion of total sequence under selection (Fig. 2a) and non-protein-coding sequence under selection (Fig. 2b) to the proportion of coding sequence for each human chromosome. Chromosome 18 contains a low overall proportion of sequence under selection, but

this is almost entirely explained by its low coding density, as there is no deficit in non-protein-coding sequence under selection. Approximately 4.2% of the bases on chromosome 18 appear to be under purifying selection, consisting of 0.6% in exons of protein-coding genes and 3.6% in non-protein-coding elements. The proportion of non-protein-coding sequence under selection is typical for human chromosomes. (Note that chromosomes 19 and 22 are outliers in this analysis; the many local gene family expansions make it difficult to assign orthology.)

As chromosomes vary widely in size, we repeated the analysis for 5-Mb windows across the human genome (Fig. 2c, d). Although there is more scatter in the data, the overall conclusion is very similar. Notably, the average proportion of non-protein-coding selected sequence in a window is ~3.8%, and is slightly negatively correlated ($R^2 = 0.08$) with the proportion of coding sequence in the window.

Our analysis shows that the density of conserved non-protein-coding sequences is largely independent of the density of protein-coding genes. It is interesting to note that examination of non-coding aligned sequences between human and chicken²³ showed a negative correlation with coding content, and a study of highly conserved non-coding sequences in intergenic regions of human chromosome 21 did not identify tight coupling to the starts and ends of genes^{24,25}.

What is the nature of the non-protein-coding elements? First, the elements might encode transcripts that are not translated into proteins, such as small RNA genes or large regulatory RNAs²⁶. Second, they might serve a structural role, with a constant density of such elements required to maintain chromosome structure independent of gene density. Such structural elements could be evolutionarily essential for maintenance of a region, but might be dispensable if the entire region were to be deleted; this might explain the recent observation in mouse that a 1-Mb deletion in a gene desert containing highly conserved elements has no discernable phenotypic effect²⁷. Third, the elements may be largely related to the regulation of protein-coding genes, but their distribution may be inversely correlated with gene density^{28,29}. It is possible that genes in gene-poor regions tend to have more elaborate regulatory controls, and this could partially explain the relative sparsity of genes in such regions. In any case, it is clear that the finished sequence of the human genome will reveal many features of biological function and provide a firm foundation for future systematic analyses.

METHODS

Generation of the gene catalogue. We started by aligning all available human RefSeq, MGC and GenBank messenger RNA sequences, as well as GenPept sequences from several species, to the finished sequence. Gene models were inspected manually to ensure accurate transcriptional start and stop sites, and to correct splice sites. Non-canonical splice sites were used only if supported by sufficient complementary DNA-based evidence. Partial transcripts (those containing a partial open reading frame (ORF) or overlapping non-coding exons of sibling transcripts) were annotated in cases for which there was firm evidence of their existence. Gene symbols for biologically characterized loci were assigned by the HUGO Gene Nomenclature Committee. See Supplementary Table S10 for a complete list of gene symbols. Our annotations are available from the Vertebrate Genome Annotation database (VEGA, http://vega.sanger.ac.uk/Homo_sapiens).

Comparative analysis: creation of synteny maps. We performed full genomic alignments of repeat masked sequence from mouse⁴ (builds 31 and 33), rat²¹ and dog (CanFam 1.0; K. Lindblad-Toh, personal communication) with the human genome sequence using the PatternHunter program³⁰. We did this for human build 34 with the Broad finished chromosomes (8, 15, 17, 18) inserted, and also for human build 35 (mouse build 31 was used against human build 34, and mouse build 33 against human build 35). From these alignments we identified collinear clusters of conserved microsynteny, which were then used to form larger syntenic segments in a hierarchical fashion. Syntenic maps and their underlying syntenic anchors serve as the basis for identification of conserved elements.

Comparative analysis: identification of conserved elements. Starting with large-scale syntenic blocks defined by the human–mouse and human–dog syntenic maps, we generated pair-wise alignments within these syntenic blocks using the PatternHunter program³⁰. We then scanned 50-bp windows with 5-bp

offset and calculated the fraction of aligning bases that were matches (discarding windows with fewer than 20 aligning bases). These percentage conservation values were locally normalized to the average conservation in the surrounding 5 Mb to generate Z-scores measuring divergence from the local average (0) for every window. We examined the joint empirical distribution of mouse and dog Z-scores for windows contained within ancestral repeat sequence (undergoing neutral evolution and believed to predate the mouse–human split) and windows overlapping coding exons (Supplementary Fig. S4a). Coding sequence is defined as all bases that are annotated as coding in any transcript. All analysis presented uses Ensembl³¹ genes on human build 35; analysis with both Ensembl and Broad annotations on build 34 yields substantially similar results (Supplementary Information).

We combined dog and mouse Z-scores to generate a ‘composite’ Z-score (see Supplementary Information). We estimated the distribution of composite Z-scores for selected sequence by decomposing the global distribution of Z-scores into two components: a ‘neutral distribution’ centred at zero and corresponding to the conservation scores for ancestral repeat sequences, and a ‘selected distribution’ consisting of the residual after subtraction of the neutral distribution (Supplementary Fig. S4b). Taking into account the relative fractions of the aligning windows in each distribution, we were able to assign a probability that a window at a given score is under purifying selection.

We then divided the genome into non-overlapping 5-Mb windows. Within each such window, we counted the number of syntenic bases, the number of syntenic 50-bp windows, and the number of 50-bp windows under selection. The fraction of coding sequence (the explanatory variable in all regressions) was taken as the number of syntenic bases annotated as coding divided by the number of syntenic bases. The fraction under selection was calculated as the sum of all selection probabilities for all windows divided by the number of syntenic windows. If windows of only a certain class were considered, the probabilities were calculated only for windows in that class. We note that, on average, windows contained within coding exons scored only slightly higher than 0.67 probability of selection, owing to the large prior probability of neutrality. Thus, the slopes of all regressions are <1. For all analyses, we discarded any 5-Mb window with less than 4 Mb of syntenically aligned sequence (retaining >85% of all windows of non-zero euchromatic length). Similar results are obtained if the discarded windows are included, but the variance is higher.

Annotation. RefSeq (release 1), mammalian gene collection (MGC, 3 February 2003), dbEST and GenBank (29 December 2002) mRNAs were aligned to the genomic assembly using BLAT³². GenPept protein sequences (3 February 2003) were aligned using BLASTX³³ and GeneWise³⁴. All gene models were created manually using these aligned sequences as evidence, following HAWK2 (www.sanger.ac.uk/Info/workshops/hawk2) transcript type conventions. Gene models derived from aligned mRNA evidence were extended when possible using spliced EST evidence at the 5′ end and spliced and unspliced EST evidence in the 3′ untranslated region (UTR). Evidence was given relative priority as follows (high–low): RefSeq/MGC, GeneWise, other mRNAs, spliced ESTs and unspliced ESTs. We found CpG islands within 2-kb upstream and 1-kb downstream of the 5′ end of 73% of known category loci, which is somewhat higher than previous reports (in the range of 61–66%; cited in ref. 3, also refs 5–7).

Received 21 January; accepted 27 June 2005.

- Hernandez, D. & Fisher, E. M. Mouse autosomal trisomy: two's company, three's a crowd. *Trends Genet.* **15**, 241–247 (1999).
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
- Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Grimwood, J. *et al.* The DNA sequence and biology of human chromosome 19. *Nature* **428**, 529–535 (2004).
- Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 10. *Nature* **429**, 375–381 (2004).
- Martin, J. *et al.* The sequence and analysis of duplication-rich human chromosome 16. *Nature* **432**, 988–994 (2004).
- Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
- Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
- Schuler, G. D. *et al.* A gene map of the human genome. *Science* **274**, 540–546 (1996).
- Schmutz, J. *et al.* Quality assessment of the human genome sequence. *Nature* **429**, 365–368 (2004).

- Yelin, R. *et al.* Widespread occurrence of antisense transcription in the human genome. *Nature Biotechnol.* **21**, 379–386 (2003).
- Veeramachaneni, V., Makalowski, W., Galdzicki, M., Sood, R. & Makalowska, I. Mammalian overlapping genes: the comparative perspective. *Genome Res.* **14**, 280–286 (2004).
- Mouchiroud, D. *et al.* The distribution of genes in the human genome. *Gene* **100**, 181–187 (1991).
- Rebhan, M. *et al.* GeneCards: encyclopedia for genes, proteins and diseases. (Weizmann Institute of Science, Bioinformatics Unit and Genome Center, Rehovot, Israel) (<http://bioinformatics.weizmann.ac.il/cards/>) (1997).
- The Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005).
- Dennehey, B. K., Gutchess, D. G., McConkey, E. H. & Krauter, K. S. Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. *Genomics* **83**, 493–501 (2004).
- Goidts, V., Szamalek, J. M., Hameister, H. & Kehler-Sawatzki, H. Segmental duplication associated with the human-specific inversion of chromosome 18: a further example of the impact of segmental duplications on karyotype and genome evolution in primates. *Hum. Genet.* **115**, 116–122 (2004).
- King, M. *Species Evolution* 72–91 (Cambridge Univ. Press, Cambridge, 1993).
- Delneri, D. *et al.* Engineering evolution to study speciation in yeasts. *Nature* **422**, 68–72 (2003).
- Gibbs, R. A. *et al.* Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
- Chiaromonte, F. *et al.* The share of human genomic DNA under selection estimated from human-mouse genomic alignments. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 245–254 (2003).
- Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
- Dermitzakis, E. T. *et al.* Comparison of human chromosome 21 conserved nongenic sequences (CNGs) with the mouse and dog genomes shows that their selective constraint is independent of their genic environment. *Genome Res.* **14**, 852–859 (2004).
- Dermitzakis, E. T., Raymond, A. & Antonarakis, S. Conserved non-genic sequences—an unexpected feature of mammalian genomes. *Nature Rev. Genet.* **6**, 151–157 (2005).
- Rastegar, M. *et al.* Sequential histone modifications at Hoxd4 regulatory regions distinguish anterior from posterior embryonic compartments. *Mol. Cell Biol.* **24**, 8090–8103 (2004).
- Nobrega, M. A., Zhu, Y., Plajzer-Frick, I., Afzal, V. & Rubin, E. M. Megabase deletions of gene deserts result in viable mice. *Nature* **431**, 988–993 (2004).
- Ovcharenko, I. *et al.* Evolution and functional classification of vertebrate gene deserts. *Genome Res.* **15**, 137–145 (2005).
- Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for long-range enhancers. *Science* **302**, 413 (2003).
- Ma, B., Tromp, J. & Li, M. PatternHunter: faster and more sensitive homology search. *Bioinformatics* **18**, 440–445 (2002).
- Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res.* **30**, 38–41 (2002).
- Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
- Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
- Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Special thanks are due to L. Gaffney for help with the manuscript, figures and tables, and to K. Lance for help with the manuscript. We are grateful to E. Eichler and X. She for sharing data on segmental duplications, T. Furey for help with lists of genetic markers and placement of RefSeqs, and K. Lindblad-Toh for sharing data from the dog genome project. In addition, we thank Ming Li and Bin Ma (Bioinformatics Solutions Inc.) for providing PatternHunter and advice about how to choose appropriate parameters. We also acknowledge the HUGO Gene Nomenclature Committee (S. Povey, E. A. Bruford, V. K. Khodiyar, R. C. Lovering, M. J. Lush, T. P. Sneddon, C. C. Talbot Jr and M. W. Wright) for assigning official gene symbols. We are grateful to all the members, present and past, of the Broad (and Whitehead) sequencing platform for the consistent high quality of their data.

Author Information Accession numbers for all clones contributing to the finished sequence of human chromosome 18 can be found in Supplementary Table S3. The updated human chromosome 18 sequence can be accessed through GenBank accession number NC_000018. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.N. (chad@broad.mit.edu).

LETTERS

A role for lateral hypothalamic orexin neurons in reward seeking

Glenda C. Harris¹, Mathieu Wimmer¹ & Gary Aston-Jones¹

The lateral hypothalamus is a brain region historically implicated in reward and motivation^{1–4}, but the identity of the neurotransmitters involved are unknown. The orexins (or hypocretins) are neuropeptides recently identified as neurotransmitters in lateral hypothalamus neurons^{5,6}. Although knockout and transgenic overexpression studies have implicated orexin neurons in arousal and sleep⁷, these cells also project to reward-associated brain regions, including the nucleus accumbens and ventral tegmental area^{8,9}. This indicates a possible role for these neurons in reward function and motivation^{3,10}, consistent with previous studies implicating these neurons in feeding⁶. Here we show that activation of lateral hypothalamus orexin neurons is strongly linked to preferences for cues associated with drug and food reward. In addition, we show that chemical activation of lateral hypothalamus orexin neurons reinstates an extinguished drug-seeking behaviour. This reinstatement effect was completely blocked by prior administration of an orexin A antagonist. Moreover, administration of the orexin A peptide directly into the ventral tegmental area also reinstated drug-seeking. These data reveal a new role for lateral hypothalamus orexin neurons in reward-seeking, drug relapse and addiction.

We used a two-chamber, nonbiased, conditioned place-preference (CCP) model to measure the rewarding properties of morphine, cocaine or food^{11–13}. In this model, one chamber becomes associated with drug or food reward through repeated pairings, whereas the other chamber is associated with no reward. Preference for reward is measured by the amount of time animals spend in the reward-associated chamber minus the time it spends in the non-rewarded chamber, when given free access to both chambers after conditioning.

Orexin-expressing neurons are located in three contiguous hypothalamic regions: lateral hypothalamus, perifornical area (PFA) and dorsomedial hypothalamus (DMH)¹⁴. To determine whether orexin neurons were stimulated during the expression of preference for different rewards, we used double-label immunohistochemistry for both orexin and the immediate early gene protein, Fos (a marker of neuronal stimulation¹⁵). Conditioned animals displayed reward-seeking by spending significantly more time in the reward-paired chamber than non-conditioned animals (Fig. 1a). Only conditioned animals that exhibited a preference for the reward-paired chamber showed increased Fos activation in lateral hypothalamus orexin cells. Significant percentages (48–52%) of orexin neurons in the lateral hypothalamus were Fos-activated in these animals after preference testing for morphine, cocaine or food reward, when compared to non-conditioned animals (17%, $P < 0.01$) (Table 1; Fig. 1b, see also Supplementary Fig. 1). The percentage of Fos activation in lateral hypothalamus orexin neurons in non-conditioned animals was not statistically different from that found in naïve untreated animals (15%, $P = 0.10$). Our findings in naïve animals are similar to baseline levels of Fos activation reported previously in awake rats¹⁶.

Enhanced activation of lateral hypothalamus orexin neurons after conditioning indicates a potential involvement of these neurons in the preference for reward-related cues. In support of this idea, the amount of Fos activation in these neurons was correlated with the intensity of reward seeking. Significant positive correlations were found between the percentages of orexin neurons that were Fos-positive in the lateral hypothalamus and preferences shown by corresponding conditioned animals ($P < 0.01$; Table 1). Thus, as preference scores increased, percentages of lateral hypothalamus orexin neurons that were Fos-activated also increased proportionally, for all three rewards. In contrast, numbers of Fos-activated non-orexin neurons in the lateral hypothalamus did not correlate with preference. Moreover, despite the fact that substantial numbers of Fos-positive orexin neurons were found in PFA and DMH, no correlations were found in these areas between preference scores and the percentage of Fos-positive orexin neurons ($P > 0.20$; Table 1). This indicates that the relationship between reward-seeking and neuronal stimulation in the hypothalamus occurs specifically in

Table 1 | Orexin and Fos double labelling

Groups	Cell types	Percentage Fos-positive	Correlations R
Morphine-conditioned $n = 12$	Orx LH	48 ± 2*	0.72, $P < 0.01^*$
	NonOrx LH	55 ± 6	
	Orx PFA	62 ± 2	
	Orx DMH	67 ± 4	
Food-conditioned $n = 8$	Orx LH	50 ± 3*	0.87, $P < .01^*$
	NonOrx LH	47 ± 5	
	Orx PFA	42 ± 3	
	Orx DMH	47 ± 6	
Cocaine-conditioned $n = 8$	Orx LH	52 ± 5*	0.90, $P < 0.01^*$
	NonOrx LH	78 ± 7	
	Orx PFA	67 ± 3	
	Orx DMH	74 ± 3	
Non-conditioned $n = 15$	Orx LH	17 ± 2	
	NonOrx LH	43 ± 6	
	Orx PFA	52 ± 4	
	Orx DMH	59 ± 4	
Naïve $n = 6$	Orx LH	15 ± 1	
	NonOrx LH	29 ± 8	
	Orx PFA	52 ± 3	
	Orx DMH	57 ± 6	
Novelty-conditioned $n = 6$	Orx LH	18 ± 2	
	NonOrx LH	50 ± 1	
	Orx PFA	56 ± 3	
	Orx DMH	63 ± 5	

The percentages of orexin-positive cells that were also Fos-positive are indicated. Correlation coefficients for the comparisons between these percentages and the corresponding preference score in each animal. Abbreviations used: Orx, orexin-positive neurons; NonOrx, orexin-negative neurons; LH, lateral hypothalamus; PFA, perifornical area; and DMH, dorsomedial hypothalamus. The non-orexin Fos-positive neurons in the lateral hypothalamus are given as total counts not percentages. Lateral hypothalamus by group ANOVA $F_{3,39} = 33$, $P < 0.01$.

*Significantly different from other groups, $P < 0.05$.

¹Laboratory of Neuromodulation and Behavior, Department of Psychiatry, University of Pennsylvania, 705 Stellar Chance/6100 422 Curie Blvd, Philadelphia, Pennsylvania 19104-6100, USA.

orexin neurons, but not in non-orexin neurons or in orexin neurons in adjacent hypothalamic areas.

A separate group of animals ($n = 11$) was given morphine CPP training followed by a systemic injection of the selective orexin A antagonist (SB 334867; ref. 17) on a second consecutive day of CPP testing. Administration of the orexin antagonist produced a significant reduction in preference (206 ± 23 s versus 87 ± 30 s; $t_{12} = 2.3$; $P < 0.05$) that was not found in a different group of animals given a vehicle injection instead (182 ± 21 s versus 145 ± 27 s; $t_{12} = 0.9$; $P = 0.39$). There were no differences in the original preferences for morphine between these two groups ($P = 0.46$). These data indicate that orexin has a role in the amount of preference expressed.

Not all reward-seeking behaviours were associated with enhanced stimulation of lateral hypothalamus orexin neurons. In a separate group of animals conditioned with novel object reward¹⁸, we found no enhanced Fos activation in lateral hypothalamus orexin neurons. In this model, exposure to a novel object is paired with one conditioning chamber, whereas no object is paired with the other side. This model produces a robust preference for the chamber paired with the novel object ($n = 6$ rats; preference score, 228 ± 34 s), and similar activity levels as in the drug and food conditioned animals on the test day. Despite this robust preference there was no significant Fos induction in the lateral hypothalamus orexin neurons during preference testing ($18 \pm 2\%$ of orexin neurons were Fos positive; Table 1; see also Supplementary Fig. 2). Measures of orexin versus non-orexin neurons that were Fos-activated were not statistically different from non-conditioned animals ($P > 0.20$) and there was no correlation with preference scores (Table 1). Novelty conditioning entailed more conditioning days than food or drug conditioning; however, it seems unlikely that these extra conditioning days

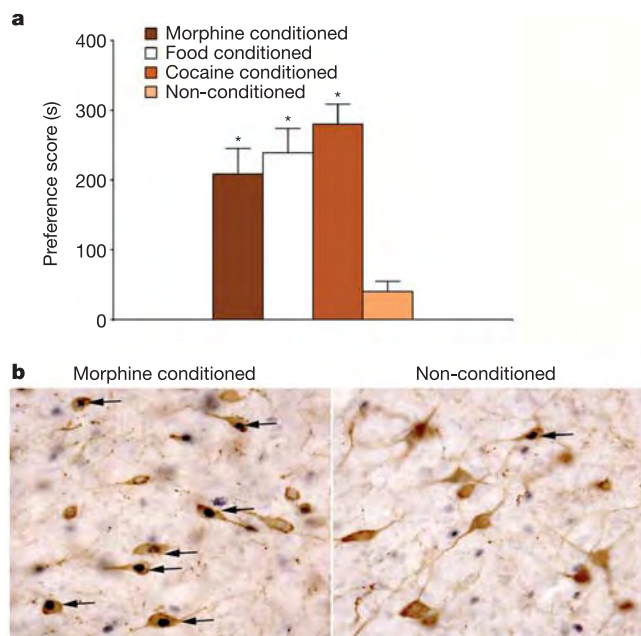


Figure 1 | Morphine conditioned animals had significantly greater place preferences and Fos activated orexin neurons in the lateral hypothalamus than non-conditioned animals. **a**, Conditioned animals showed significant preferences for the reward-paired chamber compared with non-conditioned animals ($F_{3,39} = 15$, $P < 0.01$). Preference scores for the morphine (8 mg kg^{-1}), cocaine (15 mg kg^{-1}) and food (lucky charms cereal)-paired environments expressed as the time spent in the reward-paired side minus the time spent on the non-rewarded side on the test day (mean $\pm$ s.e.m.). **b**, High-power photomicrograph of the lateral hypothalamus showing the double labelling of orexin (brown cytoplasm) and Fos protein (black nuclei) in morphine-conditioned and non-conditioned animals. Black arrows indicate double-labelled cells.

decreased orexin neuron stimulation. These data indicate that not all rewards work through the activation of the lateral hypothalamus orexin system, and that this system may be specifically engaged with consummatory rewards (that is, food and drugs).

These results show that lateral hypothalamus orexin neurons are activated in proportion to preference for certain rewards. We reasoned that if activation of these neurons drives reward seeking, then stimulation of these cells should reinstate extinguished preference. To test this, we submitted rats to morphine place conditioning and then extinguished the morphine preference by repeatedly exposing the animals to the chambers without morphine administration¹⁹ (Fig. 2a). To activate orexin neurons, we microinfused the Y4 agonist rPP (rat pancreatic polypeptide) directly into the lateral hypothalamus orexin neuronal area. We expected rPP to stimulate orexin neurons because these cells express the Y4 receptor, and activation of this receptor by rPP potentially induces Fos activation in lateral hypothalamus orexin neurons²⁰. We found that micro-injecting rPP into the lateral hypothalamus robustly reinstated an extinguished morphine place preference (Fig. 2a). Similar control injections of rPP dorsal, ventral or medial to lateral hypothalamus orexin cells did not reinstate preference (Fig. 2a, b). To insure that vehicle injections alone did not produce reinstatement, five animals from the lateral hypothalamus-injected group were given vehicle injections 3 days before the rPP injection. In these cases, the vehicle

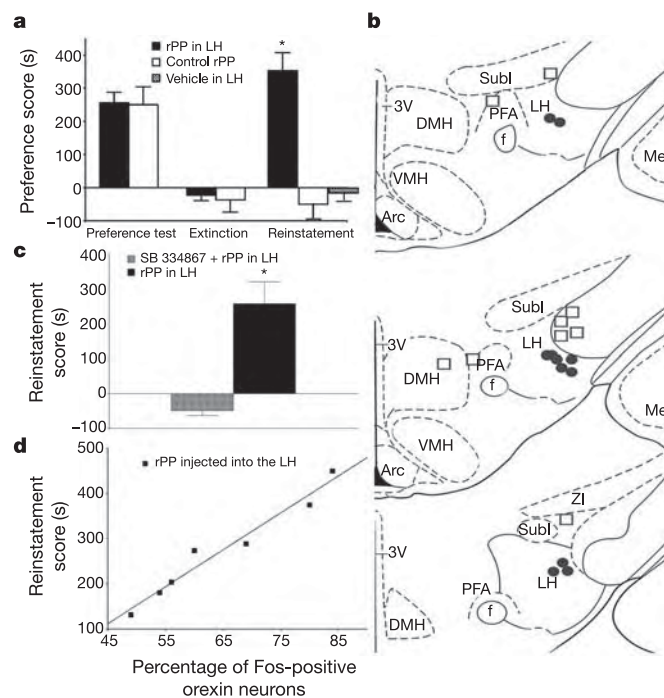


Figure 2 | Activation of lateral hypothalamus orexin neurons by rPP reinstated an extinguished preference for morphine. **a**, Preference scores are shown for both rPP- (150 nM) and vehicle-injected groups (mean $\pm$ s.e.m. in morphine-paired side minus saline-paired side) during the initial conditioning test, after extinction and during the reinstatement test. **b**, Effective (filled circles, $n = 12$) and ineffective (open squares, $n = 9$) sites of rPP administration. **c**, The selective orexin A antagonist, SB 334867 ($20\text{--}30 \text{ mg kg}^{-1}$), blocked reinstatement by rPP ($n = 8$). Data were included only if rPP injection into lateral hypothalamus on the following day (without the antagonist pretreatment) produced reinstatement of preference. **d**, Plot of correlation between reinstatement scores and percentages of lateral hypothalamus orexin neurons that were Fos activated in rPP reinstated animals. Abbreviations used were: ZI, zona incerta; Sub I, subincertal nucleus; Me, medial amygdala nucleus; VMH, ventromedial hypothalamic nucleus; Arc, arcuate nucleus; PFA, perifornical area; DMH, dorsomedial hypothalamus; LH lateral hypothalamus; 3V, third ventricle; and f, fornix.

injections did not reinstate the CPP behaviour, whereas the later rPP injections produced robust reinstatement (Fig. 2a; reinstatement score: -47 ± 71 s for vehicle versus 465 ± 100 s for rPP). Systemic injections of morphine are a reliable method to reinstate extinguished responses to morphine cues in this model¹⁹. We found that the reinstatement by rPP in the lateral hypothalamus was similar to the reinstatement produced by systemic morphine (8 mg kg^{-1} , intraperitoneal; morphine reinstatement score was 424 ± 103 s, $n = 8$ rats and rPP reinstatement score was 353 ± 52 s, $n = 12$, $P > 0.5$).

The rPP-induced reinstatement was completely blocked by prior systemic administration of the orexin antagonist, SB 334867 (Fig. 2c), confirming a role for orexin receptors in rPP-induced reinstatement. One day later, the same animals reinstated preference when injected with rPP in the lateral hypothalamus without antagonist pretreatment. This confirmed that antagonist pretreatment blocked an otherwise effective reinstatement by rPP. In addition, we measured Fos staining in both orexin and non-orexin lateral hypothalamus cells in seven rats with significant reinstatement after rPP injections, and in six dorsal control animals that did not reinstate (reinstatement scores were 279 ± 40 s and -51 ± 46 s, respectively; $P < 0.01$). In the animals for which rPP reinstated preference, the percentage of Fos-activated orexin cells in the lateral hypothalamus was significantly greater than in control animals, for which rPP did not cause reinstatement ($P < 0.01$; lateral hypothalamus orexin neurons that were Fos-activated for reinstated rats, $66 \pm 8\%$; and for control non-reinstated animals, $11 \pm 5\%$). Moreover, a highly significant correlation was found between the preference expressed during rPP-induced reinstatement and the percentage of lateral hypothalamus orexin neurons that were Fos-activated (Fig. 2d; $R = 0.99$; $P > 0.01$). Thus, the greater the number of lateral hypothalamus orexin cells that were stimulated, the greater the intensity of reinstatement that resulted. No differences were found in the number of non-orexin Fos-positive cells between the groups ($P = 0.77$), and no correlations were found between the numbers of these non-orexin Fos-activated neurons and reinstatement scores ($R = 0.23$; $P = 0.72$).

The above evidence strongly indicates that stimulation of orexin lateral hypothalamus neurons can drive reinstatement behaviour for cues associated with drug reward. The effect of rPP on morphine

reinstatement could be due to a stress/arousal effect, or to the fact that rPP mimics the motivational effects of morphine. Our preliminary results indicate that an acute injection of morphine (but not saline) in the CPP box strongly induces Fos in lateral hypothalamus orexin neurons, whereas, exposure to footshock (using parameters that we and others find reinstates morphine CPP)¹⁹ only activated orexin neurons in the DMH and PFA but not in the lateral hypothalamus. These data support the latter interpretation (see Supplementary Table 1).

To determine whether the orexin peptide could produce reinstatement of morphine reward seeking we microinjected orexin A into the ventral tegmental area (VTA). The VTA contributes to many forms of drug reinstatement²¹ and orexin has an excitatory action on GABA- and dopamine-containing neurons in this area²². We found that orexin caused a significant reinstatement response for morphine reward only when microinjected into VTA, but not when injected into areas surrounding VTA (Fig. 3a, b). Similar vehicle injections had no effect.

Taken together these data strongly indicate that orexin neurons in the lateral hypothalamus become activated by cues associated with consummatory rewards such as food and drugs. This view is consistent with an early proposal for orexin function in feeding⁶ and recent implications in morphine dependence and withdrawal²³. Orexin receptors are expressed at high levels in reward-associated areas, including the VTA and nucleus accumbens^{24,25}. As both food and drugs of abuse work through common brain substrates^{10,26}, it is not surprising that lateral hypothalamus orexin neurons may mediate responses to cues associated with both food and drugs, but not non-consummatory rewards (for example, novelty). This is consistent with a recent report showing that treatments which increase consummatory behaviour were associated with Fos activation of lateral hypothalamus orexin cells more than exposure to novelty²⁷. Our study also indicates that different sets of hypothalamic orexin neurons may have different functions. A previous report indicated that Fos activation of orexin neurons in PFA and DMH showed diurnal changes consistent with a role in production or maintenance of arousal; lateral hypothalamus orexin neurons showed no such diurnal property²⁸. These results, together with our findings, suggest the hypothesis that orexin neurons in PFA and DMH regulate arousal whereas those in lateral hypothalamus regulate reward processing. Our data further indicate that activation of lateral hypothalamus orexin neurons may be important in the reinstatement of drug-seeking behaviours and thereby have an important role in modulating rewarding behaviour and addiction. Although orexin administration directly into VTA reinstated morphine preference, VTA may not be the only location for such orexin effects as orexin-containing fibres are found throughout areas associated with reward processing, such as the extended amygdala¹⁰ and mesocorticolimbic system⁹. We conclude that the lateral hypothalamus orexin system is an integral part of circuitry that integrates environmental cues with consummatory rewards, and that stimulation of these neurons can drive relapse of drug-seeking behaviour.

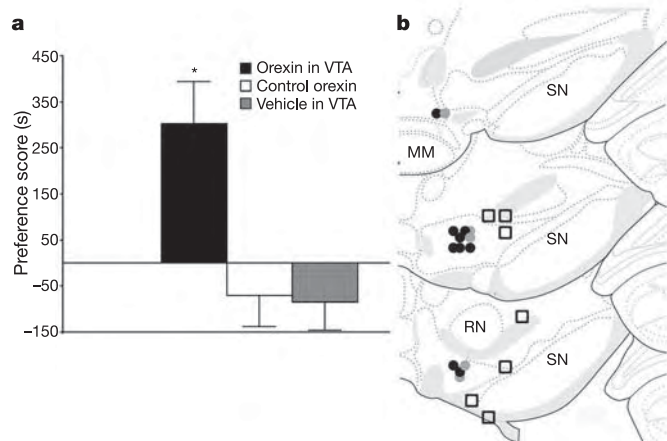


Figure 3 | Orexin administration into the VTA reinstated an extinguished preference for morphine. **a**, Reinstatement scores (time is mean $\pm$ s.e.m. in morphine-paired side minus saline-paired side) after orexin administration (140 nM) in the VTA. Orexin in the VTA ($n = 8$) was significantly different from vehicle in the VTA ($n = 6$) or orexin outside the VTA ($n = 7$) in reinstating an extinguished morphine preference ($F_{2,18} = 11$, $P < 0.01$). **b**, Effective (filled circles) and ineffective (open squares) sites of orexin administration. Gray circles are sites of vehicle administration. Drawings were adapted from ref. 31 (plates 36–38). Abbreviations used were: SN, substantia nigra; MM, medial mammillary nucleus; and RN, red nucleus.

METHODS

Subjects. Male Sprague–Dawley rats (200–250 g; Harlan) were group-housed in accordance with NIH guidelines on a 12-h light/dark cycle with food and water available *ad libitum*. The Institutional Animal Care and Use Committee of the University of Pennsylvania approved all animal procedures.

Drugs. Morphine sulphate powder, provided by the National Institute on Drug Abuse, was dissolved in sterile saline and administered via intraperitoneal injection (8 mg kg^{-1}). Cocaine (Sigma-Aldrich) was dissolved in sterile saline (15 mg kg^{-1} , intraperitoneal). SB-334867 (Tocris) was dissolved in a 10% (w/v) encapsin/2% DMSO solution in water (intraperitoneal, 20 mg kg^{-1} ($n = 2$) and 30 mg kg^{-1} ($n = 18$); refs 29, 30). Orexin A (140 nM ; Tocris) and rPP (150 nM^{20} ; Anaspec) were dissolved in artificial cerebral spinal fluid (vehicle).

Surgery and histology. Rats were anaesthetized with a ketamine/xylazine cocktail and implanted with bilateral chronic indwelling guide cannulae

aimed 2 mm above the lateral hypothalamus (anterior/posterior (AP) -3.3 , medial/lateral (ML) $+2.5$ and dorsal/ventral (DV) -8.0 , 5° angle from bregma) or VTA (AP -5.3 , ML $+2.5$, DV -7.5). Surgeries were similar to previous reports¹². Control injections were made ~ 2 mm above the lateral hypothalamus or VTA. Animals were given 1 week to recover from surgery before place conditioning began. After each experiment, animals were killed with an overdose of sodium pentobarbital (100 mg kg^{-1} intraperitoneal) and cannulae locations were confirmed in histological analyses.

Conditioned place preference procedure. We used a standard two-chamber balanced design, with two conditioning sessions (for food and drugs) occurring for 3 days consecutively. Injections of drug or saline (food or no food) alternated between morning and afternoon sessions. Conditioning and testing were identical to our previously published reports for morphine¹¹, cocaine¹² and food¹³. Food conditioned animals were not food deprived and were pre-exposed to the sweet flavoured cereal one week before conditioning. Non-conditioned animals consisted of morphine-treated animals for which the environment was explicitly unpaired with morphine ($n = 7$)¹¹, and cocaine-treated animals in which conditioning failed (animals who were conditioned but whose preference scores were less than 70 s for the cocaine environment, $n = 8$). No significant differences ($P > 0.08$) were found between these two non-conditioned groups in any measures taken, and their data were pooled. Novel object conditioning¹⁸ included 10-min pairings of a novel object versus no object in distinct compartments for 8 d, with 1 h between pairings ($n = 6$). The novel objects were paper bows, a plastic ball, a piece of cotton, a newspaper, halves of socks, pieces of egg carton, a sponge and wood chips. The order of pairings (object versus no object) was alternated each day.

Extinction and reinstatement procedure. After conditioning, animals were tested daily until preference levels for the morphine-paired chamber dropped to less than 70 s difference for the two chambers for two consecutive days. rPP or vehicle was injected bilaterally into the lateral hypothalamus (or, orexin administered into the VTA) 30 min before the reinstatement test. For the orexin antagonist experiments, SB-334867 was injected 30 min before rPP. For morphine reinstatement, morphine 8 mg kg^{-1} (intraperitoneal) was injected immediately before placing animals in the conditioning boxes.

Double label immunohistochemistry. Two hours after CPP testing rats were deeply anaesthetized and perfused transcardially with paraformaldehyde as previously described¹¹. Brain sections were first processed for Fos-immunostaining with nickel ammonium sulphate intensification of 3',3'-diaminobenzidine (DAB) as we have previously described¹¹, and then processed for orexin (orexin A antibody, Santa Cruz Biotechnology; 1:1,000; biotinylated secondary, 1:500). Slices were then mounted on glass slides, dehydrated in alcohol and sealed with a cover slip. Double-labelled cells were readily identified because orexin stained the cytoplasm brown and Fos immunoreactive nuclei were black. For counting, two sections from each animal were chosen at the same levels with equivalent numbers of orexin-positive neurons. Colour photographs of the lateral hypothalamus and DMH were taken and the numbers of orexin neurons, and Fos-positive/orexin doubly labelled neurons were counted and averaged for each animal. The number of double-labelled cells were determined by both blinded and non-blinded observers; there was a 94% agreement between these two sets of observations. All orexin-labelled neurons lateral to the fornix were considered to be in the lateral hypothalamus. All orexin labelled neurons located dorsal and 0.4 mm medial to the fornix were considered to be in the PFA. All remaining orexin labelled neurons from the medial edge of the PFA region to the third ventricle, were considered to be in the DMH.

Data analyses. Place conditioning data were analysed by calculating the time spent in the reward-paired chamber minus the time spent in the other chamber. The resulting difference score was compared between groups using analysis of variance (ANOVA). The percentages of double-labelled cells were compared between groups using ANOVA. Pearson's R correlation was used to compare preference or reinstatement scores with the number of double labelled cells. Where necessary, post-hoc analysis was carried out with a Newman-Keuls test.

Received 31 May; accepted 26 July 2005.

Published online 14 August 2005.

- Anand, B. K. & Brobeck, J. R. Hypothalamic control of food intake in rats and cats. *Yale J. Biol. Med.* **24**, 123–140 (1951).
- Olds, J. & Milner, P. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. Comp. Physiol. Psychol.* **47**, 419–427 (1954).
- DiLeone, R. J., Georgescu, D. & Nestler, E. J. Lateral hypothalamic neuropeptides in reward and drug addiction. *Life Sci.* **73**, 759–768 (2003).
- Petrovich, G. D. & Gallagher, M. Amygdala subsystems and control of feeding behaviour by learned cues. *Ann NY Acad. Sci.* **985**, 251–262 (2003).
- de Lecea, L. et al. The hypocretins: hypothalamus-specific peptides with neuroexcitatory activity. *Proc. Natl Acad. Sci. USA* **95**, 322–327 (1998).
- Sakurai, T. et al. Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behaviour. *Cell* **92**, 573–585 (1998).
- Chemelli, R. M. et al. Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation. *Cell* **98**, 437–451 (1999).
- Peyron, C. et al. Neurons containing hypocretin (orexin) project to multiple neuronal systems. *J. Neurosci.* **18**, 9996–10015 (1998).
- Fadel, J. & Deutch, A. Y. Anatomical substrates of orexin-dopamine interactions: lateral hypothalamic projections to the ventral tegmental area. *Neuroscience* **111**, 379–387 (2002).
- Baldo, B. A., Daniel, R. A., Berridge, C. W. & Kelley, A. E. Overlapping distributions of orexin/hypocretin- and dopamine-beta-hydroxylase immunoreactive fibers in rat brain regions mediating arousal, motivation, and stress. *J. Comp. Neurol.* **464**, 220–237 (2003).
- Harris, G. C. & Aston-Jones, G. Enhanced morphine preference following prolonged abstinence: association with increased Fos expression in the extended amygdala. *Neuropsychopharmacology* **28**, 292–299 (2003).
- Harris, G. C. & Aston-Jones, G. Critical role for ventral tegmental glutamate in preference for a cocaine-conditioned environment. *Neuropsychopharmacology* **28**, 73–76 (2003).
- Harris, G. C. & Aston-Jones, G. Altered motivation and learning following opiate withdrawal: evidence for prolonged dysregulation of reward processing. *Neuropsychopharmacology* **28**, 865–871 (2003).
- Siegel, J. M. Hypocretin (orexin): role in normal behaviour and neuropathology. *Annu. Rev. Psychol.* **55**, 125–148 (2004).
- Herrera, D. G. & Robertson, H. A. Activation of c-fos in the brain. *Prog. Neurobiol.* **50**, 83–107 (1996).
- Espana, R. A., Valentino, R. J. & Berridge, C. W. Fos immunoreactivity in hypocretin-synthesizing and hypocretin-1 receptor-expressing neurons: effects of diurnal and nocturnal spontaneous waking, stress and hypocretin-1 administration. *Neuroscience* **121**, 201–217 (2003).
- Smart, D. et al. SB-334867-A: the first selective orexin-1 receptor antagonist. *Br. J. Pharmacol.* **132**, 1179–1182 (2001).
- Bevins, R. A. et al. Novel-object place conditioning: behavioural and dopaminergic processes in expression of novelty reward. *Behav. Brain Res.* **129**, 41–50 (2002).
- Wang, B., Luo, F., Zhang, W. T. & Han, J. S. Stress or drug priming induces reinstatement of extinguished conditioned place preference. *Neuroreport* **11**, 2781–2784 (2000).
- Campbell, R. E. et al. Orexin neurons express a functional pancreatic polypeptide Y4 receptor. *J. Neurosci.* **23**, 1487–1497 (2003).
- Shalev, U., Grimm, J. W. & Shaham, Y. Neurobiology of relapse to heroin and cocaine seeking: a review. *Pharmacol. Rev.* **54**, 1–42 (2002).
- Korotkova, T. M., Sergeeva, O. A., Eriksson, K. S., Haas, H. L. & Brown, R. E. Excitation of ventral tegmental area dopaminergic and nondopaminergic neurons by orexin/hypocretins. *J. Neurosci.* **23**, 7–11 (2003).
- Georgescu, D. et al. Involvement of the lateral hypothalamic peptide orexin in morphine dependence and withdrawal. *J. Neurosci.* **23**, 3106–3111 (2003).
- Trivedi, P., Yu, H., MacNeil, D. J., Van der Ploeg, L. H. & Guan, X. M. Distribution of orexin receptor mRNA in the rat brain. *FEBS Lett.* **438**, 71–75 (1998).
- Marcus, J. N. et al. Differential expression of orexin receptors 1 and 2 in the rat brain. *J. Comp. Neurol.* **435**, 6–25 (2001).
- Carr, K. D. Augmentation of drug reward by chronic food restriction: behavioural evidence and underlying mechanisms. *Physiol. Behav.* **76**, 353–364 (2002).
- Baldo, B. A. et al. Activation of a subpopulation of orexin/hypocretin-containing hypothalamic neurons by GABAA receptor-mediated inhibition of the nucleus accumbens shell, but not by exposure to a novel environment. *Eur. J. Neurosci.* **19**, 376–386 (2004).
- Estabrooke, I. V. et al. Fos expression in orexin neurons varies with behavioural state. *J. Neurosci.* **21**, 1656–1662 (2001).
- Rodgers, R. J. et al. SB-334867, a selective orexin-1 receptor antagonist, enhances behavioural satiety and blocks the hyperphagic effect of orexin-A in rats. *Eur. J. Neurosci.* **13**, 1444–1452 (2001).
- Haynes, A. C. et al. A selective orexin-1 receptor antagonist reduces food consumption in male and female rats. *Regul. Pept.* **96**, 45–51 (2000).
- Swanson, L. W. *Brain Maps: Structure Of The Rat Brain* (Elsevier, Amsterdam, 1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Shiekhata and R. Smith for comments on the manuscript; and Y. Zhu and S. Aston-Jones for assistance with illustrations and photography. This work was supported by NIH.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.C.H. (glenda@dolphin.upenn.edu).

LETTERS

Dependence of *Drosophila* wing imaginal disc cytonemes on Decapentaplegic

Frank Hsiung¹, Felipe-Andrés Ramirez-Weber^{1†}, D. David Iwaki¹ & Thomas B. Kornberg¹

The anterior/posterior (A/P) and dorsal/ventral (D/V) compartment borders that subdivide the wing imaginal discs of *Drosophila* third instar larvae are each associated with a developmental organizer. Decapentaplegic (Dpp), a member of the transforming growth factor- β (TGF- β) superfamily, embodies the activity of the A/P organizer. It is produced at the A/P organizer and distributes in a gradient of decreasing concentration to regulate target genes, functioning non-autonomously to regulate growth and patterning of both the anterior and posterior compartments^{1–3}. Wingless (Wg) is produced at the D/V organizer and embodies its activity^{4,5}. The mechanisms that distribute Dpp and Wg are not known, but proposed mechanisms include extracellular diffusion⁶, successive transfers between neighbouring cells^{7,8}, vesicle-mediated movement⁹, and direct transfer via cytonemes¹⁰. Cytonemes are actin-based filopodial extensions that have been found to orient towards the A/P organizer from outlying cells. Here we show that in the wing disc, cytonemes orient towards both the A/P and D/V organizers, and that their presence and orientation correlates with Dpp signalling. We also show that the Dpp receptor, Thickveins (Tkv), is present in punctae that move along cytonemes. These observations are consistent with a role for cytonemes in signal transduction.

Cytonemes appear as fluorescent strands emanating from the apical surface of disc cells that express green fluorescent protein (GFP)¹⁰. Using standard epifluorescence microscopy, cytonemes are visible only if neighbouring cells have low background fluorescence, only in unfixed discs, and only if they extend in a single optical plane. The contour of the apical surface of the notum primordium is rather flat (see Supplementary Fig. 1), and cytonemes can be imaged in this region in discs that are suspended in liquid. However, the wing pouch primordium is convex, and cytonemes can only be imaged in discs that have been slightly flattened. The fragile nature of disc cells requires that physical and osmotic insults be minimized, and the methods we have developed to image wing cytonemes avoid rupture, delamination and other responses to injury (see Methods). Typical examples of successful preparations with flattened discs are shown in Fig. 1.

In the micrograph shown in Fig. 1b, small clones (averaging 10–15 cells) express CD8–GFP and are visible in a speckled pattern. High-magnification views of clones in this and similar discs reveal cytonemes extending outwards from some, but not all clones. On the basis of the presence or absence of cytonemes and on the orientation of cytonemes, three regions of the disc can be distinguished. In the wing blade primordium, approximately 20% of the clones extended cytonemes oriented towards either the A/P (Fig. 1c, i) or D/V compartment borders (Fig. 1h, i). More than 95% of these clones had cytonemes oriented towards one of the two borders, and <5% had cytonemes orientated towards both. Figure 1 shows a clone

imaged at two adjacent optical planes: A/P-oriented cytonemes are visible in the more apical plane and D/V-oriented cytonemes are visible in the more basal plane of the clone. We have not been able to establish whether all cells extend cytonemes, whether cells can extend more than one cytoneme, or whether a single cell can extend

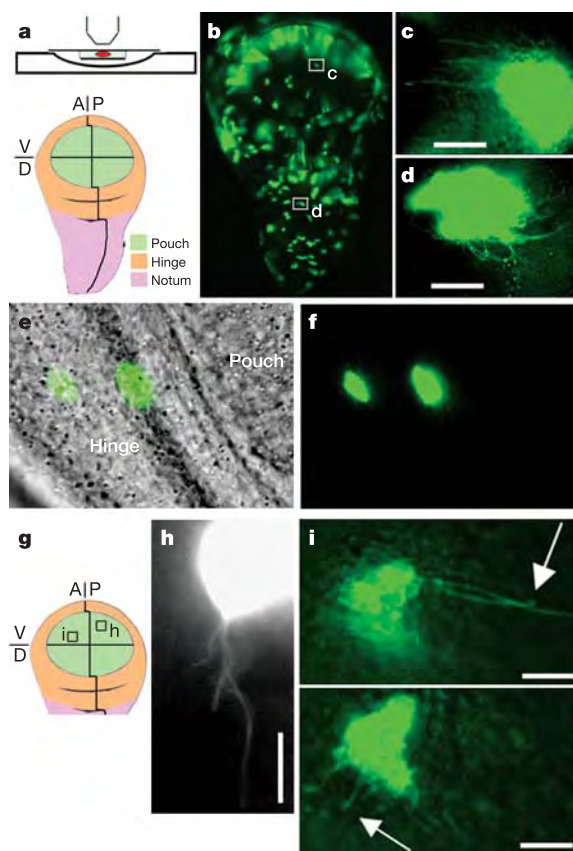


Figure 1 | Cytoneme profile. **a**, Schematic of the mounting method (top panel) and drawing of a wing disc with A/P (vertical black line) and D/V (horizontal black line) borders marked (bottom panel). **b–d**, Disc with randomly induced clones expressing CD8–GFP (**b**), with cytonemes visible at the higher magnifications (**c**, **d**). Wing pouch cytonemes are A/P-oriented whereas notum cytonemes are not. **e**, **f**, Cells labelled by clones in the hinge region do not extend cytonemes. **g**, **h**, Drawing showing a clone in the ventral-posterior region of the pouch with D/V-oriented cytonemes (**h**). **i**, Consecutive optical sections showing a clone in the ventral-anterior region of the pouch (marked in **g**) extending both A/P (top panel) and D/V (bottom panel) cytonemes. All clones are marked with CD8–GFP. Scale bars, 10 μ m.

¹Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143, USA. [†]Present address: Department of Biology, Cell and Molecular Biology Group, San Francisco State University, San Francisco, California 94132, USA.

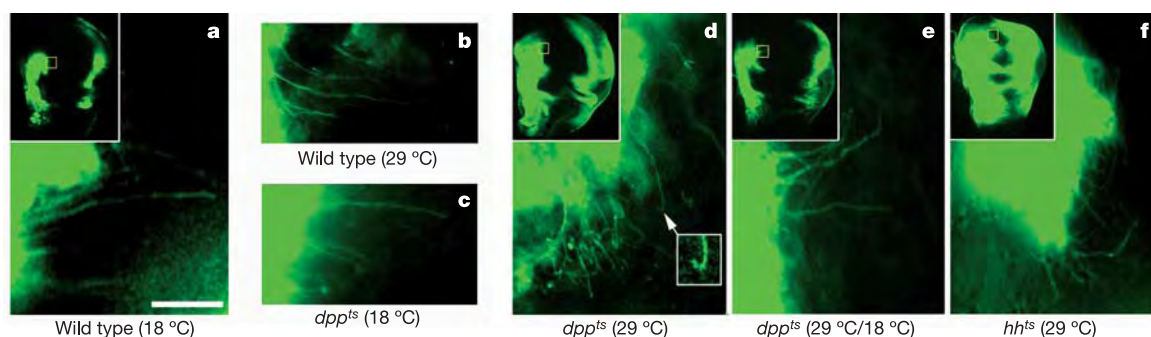


Figure 2 | Dpp maintains A/P cytoneme orientation. **a**, *brinker* expression in the wing disc (inset) and A/P-oriented cytonemes extending from *brk*-expressing cells in a wild-type background. **b**, **c**, Controls showing that the orientation and distribution of cytonemes is normal in wild-type wing discs after incubation at 29 °C (**b**) and in *dpp*^{ts} mutants at 18 °C (**c**). **d**, In *dpp*^{ts} mutant discs, the *brk* expression domain expands towards the centre after

5 h at 29 °C. Cytonemes increase in number and orient irregularly, some with 'hooked' tips (enlargement in **d**). **e**, The phenotype returns to nearly wild-type after a 3-h recovery period. **f**, *hh*^{ts} mimics the effects of *dpp*^{ts}. All images show *brk* > *CD8-GFP* in either a wild-type (**a**, **b**) or mutant background (**c**–**f**). Scale bar in **a**, 10 µm. All figures shown at the same scale.

cytonemes towards both axes. A/P cytonemes as long as 80.2 µm have been recorded; the average length in these preparations is 20.8 µm. D/V cytonemes were shorter, averaging 8.8 µm (Supplementary Table 1).

Clones in the notum primordium radiated cytonemes in all directions, without a consistent bias towards either the A/P or D/V axes of the disc. Cytonemes associated with notum clones averaged 7.4 µm in length, almost 65% shorter than A/P cytonemes in the wing primordium. Unlike cells in the wing and notum primordia, cells in the hinge/pleural primordium did not extend cytonemes. We examined these hinge/pleural cells by expressing GFP in clones (Fig. 1e) and by using an enhancer trap expressed in the hinge region (not shown), but did not observe cell extensions in either case.

Although Dpp is essential for cell survival in the notum¹¹, there is no indication that the A/P compartment border in the notum has an associated organizing centre, and it is ambiguous whether Dpp functions as a morphogen in the notum region. Dpp is apparently not required for either growth or cell survival in the hinge/pleural region¹². Thus, directional cytonemes are present in the wing primordium (where Dpp functions as a morphogen), cytonemes are present and 'omni-directional' in the notum (where Dpp may function only to support cell proliferation), and cytonemes are absent in the hinge/pleural region, where Dpp function appears not to be required. These three distinct cell types—cells with A/P- or D/V-oriented cytonemes, cells with cytonemes lacking a directional bias, or cells without cytonemes—could each be imaged in a restricted and defined region of the same disc.

We tested whether the shape and distribution of cytonemes in wing discs correlates with the presence of Dpp. First, when we reduced Dpp levels at the A/P organizer using temperature-sensitive mutants of either *dpp* (*dpp*^{ts}) or *hedgehog* (*hh*^{ts}) (*dpp* expression depends upon Hh signalling), cytonemes in the wing primordium were affected. At the permissive temperature (18 °C) in mutant discs, or at either the permissive or non-permissive (29 °C) temperatures in normal discs, cytonemes emanating from cells at the lateral flanks of discs oriented towards the A/P organizer (Fig. 2a–c). After incubation at 29 °C, however, cytonemes in *hh*^{ts} and *dpp*^{ts} mutant discs were more numerous (>twofold), and were not uniformly oriented towards the A/P organizer region (Fig. 2d, f). In these mutant discs, we observed curved and bent cytonemes, and cytonemes crossing over each other. Cytonemes with such shapes were never observed under normal conditions, or if mutant larvae were returned to the permissive temperature after a period of incubation at 29 °C (Fig. 2e).

Second, to test whether Dpp is sufficient for cytoneme induction, we imaged cytonemes in discs in which Dpp was expressed ubiquitously. In control discs, cells projected cytonemes towards A/P and D/V axes only (Fig. 3a), but in discs with heat-shock-induced Dpp (*hs-dpp*) >50% of the clones projected cytonemes outwards in all

directions (Fig. 3b). These cytonemes were significantly shorter than those in untreated discs, averaging about 10.6 µm in length. Even more striking were the cells in the hinge domain, which normally do not extend cytonemes. Under conditions of ubiquitous Dpp expression, cells in the hinge domain extended cytonemes in apparently random orientations (Fig. 3c). Our ability to image cytonemes is limited to preparations in which discs have been extracted from larvae and the cytonemes are static, but their varied appearance under the conditions we have tested illustrates that they are dynamic *in vivo*. Although we favour a model in which they extend from cells in random directions but become stabilized when functional contacts are made with signalling cells, we cannot exclude the possibility that their directionality is directly influenced by extracellular cues.

Third, we monitored the distribution of the Dpp receptor Tkv, as a Tkv–GFP fusion protein¹³. When expressed in the lateral flanks of wing discs, most of the fluorescence was localized to the plasma membrane of expressing cells. However, bright, motile punctae were also present in more central regions, as far as 30 µm from the edge of the expression domain (Fig. 4a). These punctae were motile, moving in both anterograde and retrograde directions, and some images clearly revealed their association with cytonemes (Fig. 4b, see Supplementary Video). Trafficking of these punctae was approximately 5–7 µm s^{−1} (Fig. 4c), a rate consistent with measured rates of vesicular movement on actin filaments¹⁴. The resolution of these studies could not establish whether they were inside or on cytonemes.

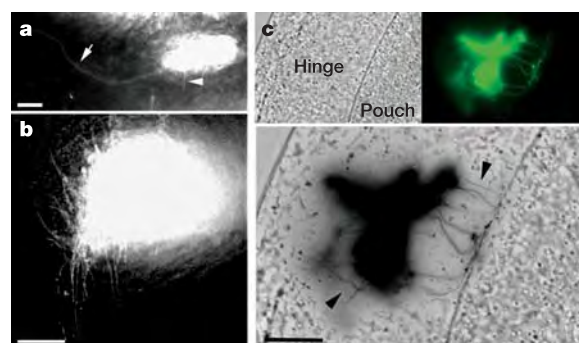


Figure 3 | Dpp overexpression induces cytoneme extensions. **a**–**c**, Cells in an *hs-dpp* background show normal A/P (arrow) and D/V (arrowhead) cytoneme orientation at 18 °C (**a**), but cells extend cytonemes in all directions after heat-shock (**b**), even in the hinge region, where cells normally do not extend cytonemes (black arrowheads in **c**, single channels shown above the merged image). Scale bars, 10 µm. Clones are marked with *CD8-GFP*.

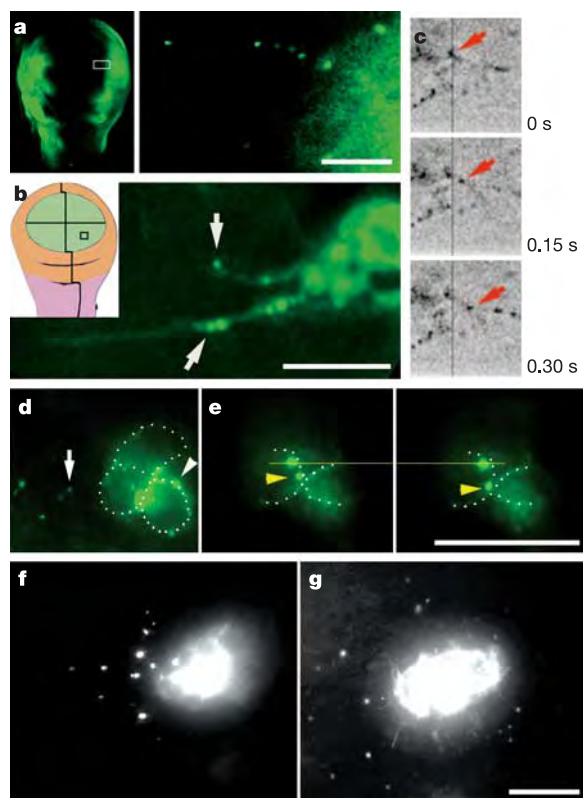


Figure 4 | Cytonemes contain motile Tkv punctae. **a**, *brk* > Tkv-GFP shows tracks of Tkv-GFP oriented towards the A/P border. **b**, Tkv-GFP punctae (arrows) can be seen localized to A/P cytonemes in Tkv-GFP clones. **c**, A motile Tkv-GFP puncta (red arrows) moving along a cytoneme. Vertical line shown as a reference point. Time shown in seconds. **d**, Control with dimethylsulphoxide shows no effect on the distribution of punctae both away from the cell bodies (arrow; cells outlined with white dots) and on the cell surface (arrowhead). **e**, Cytochalasin D reduces the number of Tkv-GFP punctae distributed at a distance from the cell bodies, but does not reduce or arrest punctae on the cell surface. Yellow arrowhead in left and right panels indicates a translocating Tkv-GFP puncta; horizontal yellow line shown as a reference point. **f**, **g**, Tkv-GFP in *hs-dpp* discs with (**f**, as in Fig. 3) and without (**g**) heat-shock. A/P border is to the left; images taken at the same scale. All scale bars, 10 μ m.

We plotted the distribution of Tkv-GFP punctae around clones in discs with normal expression of Dpp or with ubiquitous Dpp expression. In normal discs, Tkv-GFP punctae were polarized in the direction of the A/P border (Fig. 4f). In contrast, Tkv-GFP punctae in heat-shocked *hs-dpp* discs were more numerous and projected in various directions all around the circumference of the clones (Fig. 4g). As these patterns of Tkv-GFP punctae could be imaged in unflattened discs, we compared their distribution in both flattened and unflattened discs. No differences were detected between the two conditions with respect to either the total number of punctae, or to the distance from or position relative to the clones (Supplementary Table 2). This confirms that the slight flattening we use to image cytonemes does not generate or substantially alter these structures.

Our previous work has demonstrated that cytonemes bind phalloidin (a specific F-actin-binding protein) and can be labelled with an actin-GFP fusion protein, suggesting that cytonemes are actin-based¹⁰. To test whether cytonemes and the movement of Tkv-GFP punctae is actin-dependent, we treated discs containing clones of Tkv-GFP-expressing cells with cytochalasin D, an actin-binding drug. The number of Tkv-GFP punctae at a distance from the cell bodies was dramatically reduced in treated discs (reduction estimated to be >90%; Fig. 4d, e). Bright punctae were observed on the surface of cells expressing Tkv-GFP, and they appeared to move along

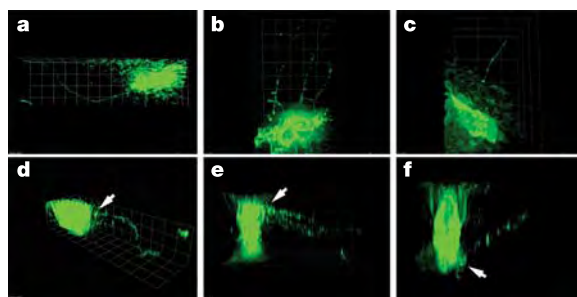


Figure 5 | Distinct apical and basal cytonemes in the wing disc. **a**, **d**, CD8-GFP-labelled clones. **b**, **e**, Tkv-GFP-labelled clones. **c**, **f**, Vav-GFP labelled clones. Top (**a**–**c**) and side (**d**–**f**) views of three-dimensional reconstructed images are shown. Apical side is up. Arrows indicate the point of cytoneme protrusions.

the surface of the cells even in the presence of drug (Fig. 4e). In contrast, the bright punctate fluorescence distant from GFP-expressing cells was not motile. These observations suggest that cytonemes can function as vehicles for active, actin-based transport of receptors.

To better document the structure of cytonemes, we reconstructed optical sections of cytoneme-producing clones to render their three-dimensional structure. In such images, we observed that cytonemes labelled with CD8-GFP (Fig. 5a, d) as well as cytonemes containing Tkv-GFP punctae (Fig. 5b, e), extend from the apical surface of the disc columnar epithelial cells. In contrast, expression of a human guanine nucleotide exchange factor (GEF) protein, Vav-GFP, which has been shown to localize to filopodia in vertebrate cells¹⁵, labels basal filopodia when expressed in wing disc cells (Fig. 5c, f). This preferential placement of proteins into different types of filopodial extensions indicates that apical and basal extensions are structurally distinct, it suggests that these cell extensions may be functionally distinct, and it implies the existence of a mechanism for sorting proteins to specific types of extensions.

Dpp is synthesized and secreted by a narrow stripe of 5–7 cells adjacent to the A/P compartment border in the wing primordium, and it distributes in a gradient of decreasing concentration that extends across the wing pouch^{8,16}. A concentration gradient does not imply a mechanism for distribution; it is conceivable that cytonemes sense and respond to Dpp but do not ferry it. However, on the basis of the results presented here, we consider cytoneme-based transport to be an attractive possibility. As this work shows, the Dpp receptor Tkv is present in cytonemes, and the presence, orientation and shape of cytonemes in wing discs correlates with what is known about the different roles that Dpp has in the wing, notum and hinge primordia. Moreover, cytonemes change in response to conditions of Dpp gain-of-function and loss-of-function. These correlations are consistent with the idea that Dpp moves from its source in an oriented manner imposed by the directionality of these cellular extensions. Several recent studies reported have cellular extensions in *Drosophila* cells that correlate with signalling by Branchless (a *Drosophila* FGF, or fibroblast growth factor)^{17–19}, Notch²⁰ and Scabrous²¹, extensions in spider cells that correlate with signalling by Dpp²², and extensions in mammalian cells that correlate with signalling by epidermal growth factor (EGF)²³. The widespread occurrence of cytonemes and cytoneme-like filopodia suggests that their role in long-distance signalling might be a general one, one that might permit selective signalling in ways that enable cells to regulate both release and uptake of signals.

METHODS

Fly stocks. *brinker (brk)-Gal4* is an enhancer trap on the X chromosome¹⁰. *UAS-CD8-GFP* was a gift from L. Luo. *dpp^{ts}* was created by placing *dpp⁴* and *dpp⁵⁶* (obtained from F. Chanut) in-trans to each other. *hh^{ts}* was obtained from the Bloomington stock centre (BL-1684), *hs-dpp* was from E. Bier. Random

CD8-GFP clones were generated by applying a heat shock to *hsFlp; abxubx > FRT > Gal4/UAS-CD8-GFP* to early third instar larvae for 12 min at 37 °C. In both *dpp^{ts}* and *hh^{ts}* experiments, larvae were grown at 18 °C, shifted to 29 °C and then immediately dissected and imaged, or were returned to the permissive temperature (18 °C) before dissection. *UAS-tkv-GFP* was created by inserting the enhanced (*E*)GFP coding sequence at the carboxy terminus of *tkv* and cloning into pUAST²⁴.

Sample preparation method. Imaginal discs were dissected and mounted in 1 × PBS buffer. Samples were placed on a coverslip, apical side down in a PBS droplet, and overlaid with a smaller coverslip to enhance flattening. The 'sandwich' was then turned upside-down, placed over the depression well of a slide and secured with halocarbon oil at the edges. Note that in the finished preparation, the samples, buffer and the smaller coverslip used for flattening, are all 'hanging' under the larger coverslip that contacts the objective lenses (see illustration in Fig. 1a). This allows the field to be scanned without damaging the sample, and minimizes additional compression of the samples. Cytochalasin D was used at 20 µg ml⁻¹ in Grace medium containing 5% dimethylsulphoxide²¹. Samples were incubated in the solution for 30 min on ice before imaging.

Imaging. Images were taken using an upright Zeiss Axioplan2 microscope equipped with a Cooke CCD camera, and were refined with deconvolution software (Slidebook v4.0, Intelligent Imaging Innovations). Tkv-GFP movement was captured with 100–153-ms exposure intervals. Time-lapse sequences and kinetic analysis of TKV-GFP trafficking were accomplished using Slidebook software. Confocal images were acquired with a Leica TCS laser scanning microscope and analysed using NIH Image J.

Received 19 October 2004; accepted 16 June 2005.

- Basler, K. & Struhl, G. Compartment boundaries and the control of *Drosophila* limb pattern by *hedgehog* protein. *Nature* **368**, 208–214 (1994).
- Tabata, T. & Kornberg, T. B. Hedgehog is a signalling protein with a key role in patterning *Drosophila* imaginal discs. *Cell* **76**, 89–102 (1994).
- Tabata, T. & Takei, Y. Morphogens, their identification and regulation. *Development* **131**, 703–712 (2004).
- Neumann, C. J. & Cohen, S. M. A hierarchy of cross-regulation involving *Notch* *wingless*, *vestigial* and *cut* organizes the dorsal/ventral axis of the *Drosophila* wing. *Development* **122**, 3477–3485 (1996).
- Zecca, M., Basler, K. & Struhl, G. Direct and long-range action of a *wingless* morphogen gradient. *Cell* **87**, 833–844 (1996).
- Strigini, M. & Cohen, S. M. *Wingless* gradient formation in the *Drosophila* wing. *Curr. Biol.* **10**, 293–300 (2000).
- Kerszberg, M. & Wolpert, L. Mechanisms for positional signalling by morphogen transport: a theoretical study. *J. Theor. Biol.* **191**, 103–114 (1998).
- Entchev, E. V., Schwabedissen, A. & Gonzalez-Gaitan, M. Gradient formation of the TGF-β homolog Dpp. *Cell* **103**, 981–991 (2000).
- Panakova, D., Sprong, H., Marois, E., Thiele, C. & Eaton, S. Lipoprotein particles are required for Hedgehog and Wingless signalling. *Nature* **435**, 58–65 (2005).
- Ramirez-Weber, F. A. & Kornberg, T. B. Cytonemes: cellular processes that project to the principal signalling center in *Drosophila* imaginal discs. *Cell* **97**, 599–607 (1999).
- Sato, M. & Saigo, K. Involvement of *pannier* and *u-shaped* in regulation of decapentaplegic-dependent *wingless* expression in developing *Drosophila* notum. *Mech. Dev.* **93**, 127–138 (2000).
- Moreno, E., Basler, K. & Morata, G. Cells compete for Decapentaplegic survival factor to prevent apoptosis in *Drosophila* wing development. *Nature* **416**, 755–759 (2002).
- Nellen, D., Burke, R., Struhl, G. & Basler, K. Direct and long-range action of a DPP morphogen gradient. *Cell* **85**, 357–368 (1996).
- Zigmond, S. H. Recent quantitative studies of actin filament turnover during cell locomotion. *Cell Motil. Cytoskeleton* **25**, 309–316 (1993).
- Kranewitter, W. J., Danninger, C. & Gimona, M. GEF at work: Vav in protruding filopodia. *Cell Motil. Cytoskeleton* **49**, 154–160 (2001).
- Teleman, A. A. & Cohen, S. M. Dpp gradient formation in the *Drosophila* wing imaginal disc. *Cell* **103**, 971–980 (2000).
- Ribeiro, C., Ebner, A. & Affolter, M. *In vivo* imaging reveals different cellular functions for FGF and Dpp signalling in tracheal branching morphogenesis. *Dev. Cell* **2**, 677–683 (2002).
- Wolf, C., Gerlach, N. & Schuh, R. *Drosophila* tracheal system formation involves FGF-dependent cell extensions contacting bridge cells. *EMBO Rep.* **3**, 563–568 (2002).
- Sato, M. & Kornberg, T. B. FGF is an essential mitogen and chemoattractant for the air sacs of the *Drosophila* tracheal system. *Dev. Cell* **3**, 195–207 (2002).
- De Jossineau, C. *et al.* Delta-promoted filopodia mediate long-range lateral inhibition in *Drosophila*. *Nature* **426**, 555–559 (2003).
- Chou, Y. H. & Chien, C. T. Scabrous controls ommatidial rotation in the *Drosophila* compound eye. *Dev. Cell* **3**, 839–850 (2002).
- Akiyama-Oda, Y. & Oda, H. Early patterning of the spider embryo: a cluster of mesenchymal cells at the cumulus produces Dpp signals received by germ disc epithelial cells. *Development* **130**, 1735–1747 (2003).
- Lidke, D. S. *et al.* Quantum dot ligands provide new insights into erbB/HER receptor-mediated signal transduction. *Nature Biotechnol.* **22**, 198–203 (2004).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We gratefully acknowledge many helpful discussions with G. Ehrenkauf, and thank K. Moses, E. Bier, L. Luo, and F. Chanut for fly stocks. This work was supported by a grant from the NIH to T.B.K.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.B.K. (tkornberg@biochem.ucsf.edu).

LETTERS

Accelerated ageing in mice deficient in Zmpste24 protease is linked to p53 signalling activation

Ignacio Varela^{1*}, Juan Cadiñanos^{1*}, Alberto M. Pendás¹, Ana Gutiérrez-Fernández¹, Alicia R. Folgueras¹, Luis M. Sánchez¹, Zhongjun Zhou^{3†}, Francisco J. Rodríguez¹, Colin L. Stewart⁴, José A. Vega², Karl Tryggvason³, José M. P. Freije¹ & Carlos López-Otín¹

Zmpste24 (also called FACE-1) is a metalloproteinase involved in the maturation of lamin A (Lmna), an essential component of the nuclear envelope^{1–3}. Both *Zmpste24*- and *Lmna*-deficient mice exhibit profound nuclear architecture abnormalities and multiple histopathological defects that phenocopy an accelerated ageing process^{1,2,4,5}. Similarly, diverse human progeroid syndromes are caused by mutations in *ZMPSTE24* or *LMNA* genes^{6–10}. To elucidate the molecular mechanisms underlying these devastating diseases, we have analysed the transcriptional alterations occurring in tissues from *Zmpste24*-deficient mice. We demonstrate that *Zmpste24* deficiency elicits a stress signalling pathway that is evidenced by a marked upregulation of p53 target genes, and accompanied by a senescence phenotype at the cellular level and accelerated ageing at the organismal level. These phenotypes are largely rescued in *Zmpste24*^{−/−}*Lmna*^{+/-} mice and partially reversed in *Zmpste24*^{−/−}*p53*^{−/−} mice. These findings provide evidence for the existence of a checkpoint response activated by the nuclear abnormalities caused by prelamin A accumulation, and support the concept that hyperactivation of the tumour suppressor p53 may cause accelerated ageing¹¹.

Alterations in nuclear envelope formation and dynamics are involved in the development of premature ageing syndromes. Thus, mice deficient in lamin A exhibit many features of premature ageing^{4,5}. Similarly, disruption of the *Zmpste24* gene encoding a metalloproteinase involved in prelamin A processing causes nuclear architecture abnormalities^{1,2}, a shortened lifespan and multiple ageing-associated phenotypes^{1,2} (Supplementary Fig. 1). The relevance of nuclear envelope alterations in accelerated ageing syndromes has been further confirmed by the finding that patients with different progeroid syndromes have mutations in *LMNA* or *ZMPSTE24* genes^{6–10}.

The availability of mice deficient in components of the lamin A–Zmpste24 system could provide new models to aid the study of mechanistic events underlying ageing in mammals. To this end, we first used oligonucleotide-based microarrays to analyse transcriptional alterations in tissues from *Zmpste24*-deficient (*Zmpste24*^{−/−}) mice. Out of 12,488 sequences present in the array and hybridized with RNA-derived probes from wild-type and knockout mouse liver, a total of 194 (1.6%) showed a higher than fivefold increase in expression levels in mutant versus control mice (Supplementary Table 1). By contrast, only 97 genes (0.8%) exhibited a greater than fivefold decrease in expression level in liver from *Zmpste24*^{−/−} mice. Analysis of genes upregulated in the liver of *Zmpste24*^{−/−} mice revealed that among the 25 genes for which expression was most upregulated, there were at least seven that have been characterized as

downstream targets of the p53 tumour suppressor¹². These p53 targets include *Gadd45a*, *p21* (also known as *Cdkn1a*), *PA26* (also known as *sestrin1*), *Btg2*, *Atf3*, *Rtp801* (also known as *Redd1* and *Ddit4*) and *Rgs16* (human synonym *A28-Rgs14*). Furthermore, in *Zmpste24*^{−/−} mice there was also upregulation of other genes not identified as direct p53 targets but also associated with the p53 signalling pathway. These genes include *Gadd45b* and *Gadd45g*, which share multiple functional features with the p53 target *Gadd45a* in their response to DNA damage and other stresses¹³. The levels of observed upregulation of p53 target genes in *Zmpste24*^{−/−} mice ranged from 31.8- to 9.1-fold for *Gadd45a* and *Btg2*, respectively. Northern blot analysis of liver RNAs from different *Zmpste24*^{−/−} mice confirmed the upregulation of all these p53-inducible genes (Fig. 1a). Similarly, *in situ* hybridization experiments and western blot analysis of p21, a representative p53 target, provided additional evidence for the induction of a p53 response in the liver of *Zmpste24*^{−/−} mice (Supplementary Fig. 2). Similar results were obtained after transcriptional profiling and northern blot analysis of heart from *Zmpste24*^{−/−} mice, an organ severely affected by the deficiency of this metalloproteinase¹ (Fig. 1a; see also Supplementary Table 1). We also detected expression differences of some p53 target genes between liver and heart from mutant mice, confirming the *in vivo* occurrence of tissue-specific induction of some p53 targets¹⁴: *Gadd45a* and *Rgs16* are upregulated in liver but not heart of *Zmpste24*^{−/−} mice (Fig. 1a), whereas *Igf1bp3* is upregulated in heart but not liver (Fig. 1b). We also observed a marked degree of variability in the levels of induction of p53 targets depending on the severity of phenotype of *Zmpste24*^{−/−} mice. Thus, mutant mice showing the most advanced progeroid signs consistently had the highest expression levels of p53 targets (Supplementary Fig. 3).

Parallel experiments performed with liver from mice deficient in lamin A, the Zmpste24 substrate, also revealed upregulation of p53 transcriptional targets (Fig. 1a; see also Supplementary Table 1). Nevertheless, despite the marked upregulation of p53 target genes in tissues from *Zmpste24*^{−/−} and *Lmna*^{−/−} mice, western blot analysis did not reveal an increase in p53 protein levels in any of them (Fig. 1c), as previously described in senescence-associated processes¹⁵. We next explored whether a putative p53 isoform present in *Zmpste24*^{−/−} cells could be responsible for the upregulation of p53 target genes, but we did not observe any alternatively spliced form of p53 in *Zmpste24*^{−/−} mice (Supplementary Table 2). Similarly, we did not observe variations in the status of p53 phosphorylation at Ser 18 and Ser 23, two post-translational changes that activate p53 function¹⁶ (Fig. 1c). We also failed to detect acetylated p53 in liver from mutant mice (data not shown). Therefore, it seems that common

¹Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Instituto Universitario de Oncología, and ²Departamento de Morfología y Biología Celular, Universidad de Oviedo, 33006 Oviedo, Spain. ³Division of Matrix Biology, Department of Biochemistry and Biophysics, Karolinska Institutet, Stockholm S-17177, Sweden. ⁴National Cancer Institute, Frederick, Maryland 21702, USA. [†]Present address: Department of Biochemistry, Faculty of Medicine, University of Hong Kong, Pok Fu Lam, Hong Kong.

*These authors contributed equally to this work.

post-translational modifications are not involved in the p53 activation mechanism probably responsible for the upregulation of several targets of this tumour suppressor in *Zmpste24*^{-/-} mice. Further studies revealed that levels of phosphorylated histone H2AX—an early marker of cell response to DNA damage—are increased in liver from *Zmpste24*^{-/-} mice (Fig. 1d), indicating the occurrence of a DNA damage response in these mice that could lead to p53 activation. These results agree with recent findings showing genomic instability in laminopathy-based premature ageing¹⁷. Taken together, the above results are consistent with the hypothesis that the nuclear abnormalities occurring in mice deficient in the lamin A–*Zmpste24* system trigger a stress response linked to the activation of a signalling pathway involving p53 transcriptional targets.

To evaluate the possibility that structural abnormalities emanating from the nuclear envelope defects observed in *Zmpste24*^{-/-} cells¹ could be sensed as p53-activating signals, we performed an ultrastructural analysis of tissues from mutant mice. This analysis revealed profound chromatin alterations in *Zmpste24*^{-/-} cells (Supplementary Fig. 4), which could evoke the genotoxic stress that elicits a p53-mediated response and leads to cellular senescence or death by apoptosis¹⁸. To analyse further the p53-linked response induced in *Zmpste24*^{-/-} mice, we examined whether cells from these mutant mice enter senescence earlier or undergo apoptosis at a higher rate than control cells. Cell senescence is associated with a spontaneous decline in growth rate and a terminal arrest of the cell cycle^{19,20}. Therefore, we first performed a comparative analysis of the proliferative ability of primary fibroblasts from these mice and their wild-type littermates. As shown in Fig. 2a, fibroblasts from 12-week-old *Zmpste24*^{-/-} mice exhibited a marked proliferative decrease when compared with control cells. This decrease was not primarily caused by an increased sensitivity to culture stress, as it was maintained under reduced (3%) oxygen conditions (Fig. 2b). To analyse the effect of *Zmpste24* deficiency on cell cycle, we

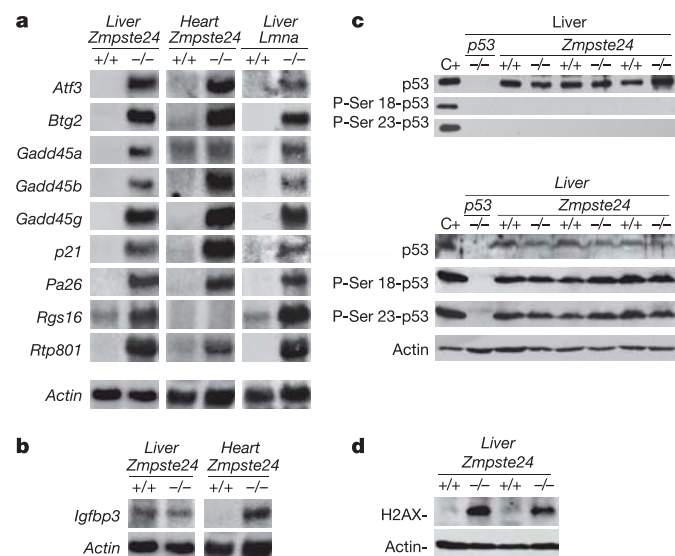


Figure 1 | Activation of a p53-induced pathway as a result of *Zmpste24* or *Lmna* deficiencies. **a**, Northern blot analysis of p53 targets in liver (left panel) and heart (middle panel) from *Zmpste24*^{-/-} mice, and in liver from *Lmna*^{-/-} mice (right panel). **b**, Northern blot showing *Igfbp3* overexpression in heart but not in liver from *Zmpste24*^{-/-} mice. **c**, Top panel: western blot analysis of p53 immunoprecipitated from *Zmpste24*^{-/-} and control livers. Doxorubicin-treated mouse fibroblasts and p53-deficient liver extracts were used as positive and negative controls, respectively. Bottom panel: western blot analysis of liver extracts from γ -irradiated *Zmpste24*^{-/-} and control mice. **d**, Western blot analysis of phosphorylated histone H2AX in liver extracts from severely affected *Zmpste24*^{-/-} mice and age-matched controls.

measured the percentage of replicative cells in fibroblast cultures after 5-bromodeoxyuridine (BrdU) exposure (Fig. 2c). Whereas 15.5% of control cells were positive, only 6.3% of the *Zmpste24*-deficient fibroblasts incorporated the nucleotide analogue. Accordingly, the percentage of cells in S phase was lower in the absence of *Zmpste24*, increasing the fraction of cells both in G1 and G2/M phases (Fig. 2d). Therefore, *Zmpste24*^{-/-} fibroblasts show a reduced proliferative activity that apparently derives from the lower replicative capacity of these cells.

In addition to growth arrest, senescence is associated with a series of distinctive molecular and morphological alterations^{19–21}. Analysis of the putative occurrence of features characteristic of senescent cells in *Zmpste24*^{-/-} mice showed that *Zmpste24*^{-/-} fibroblasts displayed flattened and enlarged morphology, and were positively stained for β -galactosidase at pH 6.0, a biological marker of senescent cells¹⁹ (Fig. 2e). Histochemical analysis of kidney sections from *Zmpste24*^{-/-} mice also revealed a strong β -galactosidase activity at pH 6.0 (Fig. 2e). We also observed a marked upregulation of secretory factors—including proteases such as cathepsin L—that may disrupt tissue integrity and function, and are associated with the development of senescence phenotypes²⁰ (Fig. 2f). Despite shortened telomere length being associated with senescence²², no significant differences in telomere length were found between wild-type and *Zmpste24*^{-/-} mice (A. Canela & M. A. Blasco, personal communication). Furthermore, cell immunostaining with annexin V and propidium iodide, or immunohistochemical analysis of activated caspase-3 levels, did not reveal major differences between

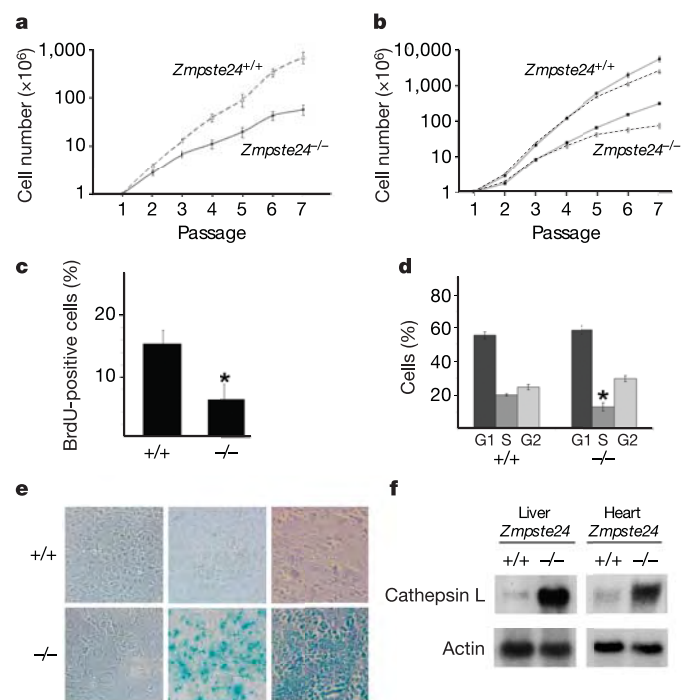


Figure 2 | Senescence in *Zmpste24*^{-/-} mice. **a**, Proliferation assay of adult fibroblasts from *Zmpste24*^{-/-} ($n = 8$) and control mice ($n = 7$). **b**, Proliferation assay of fibroblasts from three wild-type and three *Zmpste24*^{-/-} mice, cultured under atmospheric (20%, dashed lines) and reduced (3%, solid lines) oxygen concentrations. **c**, BrdU incorporation in fibroblasts of *Zmpste24*^{-/-} ($n = 2$) and control ($n = 2$) mice at passage 2. **d**, Cell cycle analysis of control ($n = 3$) and *Zmpste24*^{-/-} ($n = 3$) fibroblasts at passage 3. **e**, Flattened and enlarged morphology of *Zmpste24*^{-/-} adult fibroblasts at passage 6 (left panel). Senescence-associated β -galactosidase assay of passage 6 adult fibroblasts (middle panel) and kidneys (right panel) from control and *Zmpste24*^{-/-} mice. **f**, Northern blot analysis of cathepsin L expression in liver and heart from *Zmpste24*^{-/-} and control mice. Error bars represent s.e.m.; asterisk indicates $P < 0.05$.

samples from mutant and control mice (Supplementary Fig. 5 and data not shown). Therefore, increased apoptosis does not have a major role in the observed reduction of the proliferative activity of mutant fibroblasts. Taken together, these data suggest that a p53-linked response to the persistent nuclear abnormalities of *Zmpste24*^{-/-} mice would cause a loss of their normal cell function mainly through activation of a senescence-like programme. The progressive decline in tissue-specific functions, probably derived from the activation of this cellular programme, would then contribute to the premature ageing phenotype exhibited by these *Zmpste24*-null mice.

The above results prompted us to hypothesize that targeting either the accumulated prelamin A or the p53 signalling pathway triggered by the nuclear envelope defects of *Zmpste24*^{-/-} mice could rescue the multiple abnormalities occurring in these mice. In relation to the first possibility, it has been reported that heterozygous *Lmna*^{+/-} mice do not show apparent abnormalities⁴, which led us to speculate that generation of *Zmpste24*^{-/-}*Lmna*^{+/-} mice would decrease prelamin A levels without generating additional defects to those derived from the protease deficiency itself. Notably, *Zmpste24*^{-/-}*Lmna*^{+/-} mice exhibit a total recovery of all progeroid phenotypes observed in *Zmpste24*-null mice (Fig. 3). During preparation of this manuscript, similar findings have been reported in another strain of *Zmpste24*^{-/-} mice²³. Similarly, western blot analysis revealed a strong reduction of both prelamin A and

lamin C protein levels in *Zmpste24*^{-/-}*Lmna*^{+/-} mice (Fig. 3). This reduction is accompanied by a recovery of the normal shape in *Zmpste24*^{-/-}*Lmna*^{+/-} nuclei (Supplementary Fig. 4). Histopathological analysis also confirmed that the appreciable defects present in tissues from *Zmpste24*^{-/-} mice were absent in *Zmpste24*^{-/-}*Lmna*^{+/-} mice. Finally, no evidence of cell senescence or induction of p53 targets was found in *Zmpste24*^{-/-}*Lmna*^{+/-} mice (Fig. 3 and data not shown).

To explore the possibility that the progeroid defects of *Zmpste24*^{-/-} mice could also be rescued by targeting the abnormally activated p53 signalling pathway, we generated *Zmpste24*^{-/-}*p53*^{-/-} mice. These mice show a partial recovery of the phenotype characteristic of the *Zmpste24* deficiency and exhibit a marked gain of weight and an increased lifespan (Fig. 4). We next examined whether the absence of p53 could also result in a recovery of the cell senescence features observed in *Zmpste24*^{-/-} mice. As shown in Fig. 4, fibroblasts from *Zmpste24*^{-/-}*p53*^{-/-} mice do not show the reduced proliferative capacity observed in *Zmpste24*^{-/-} cells. Moreover, no senescence-

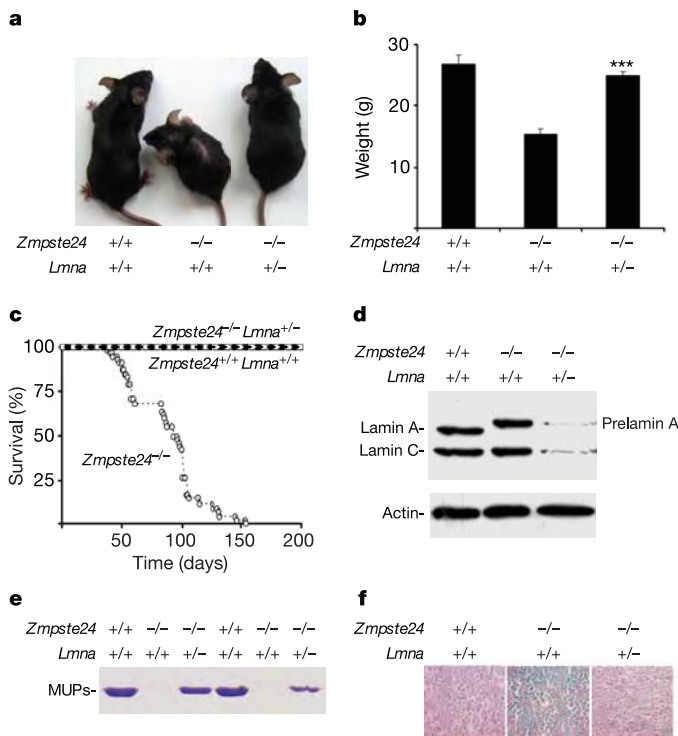


Figure 3 | *Lmna* heterozygosity rescues the *Zmpste24*^{-/-} phenotype. **a**, Photograph of 4-month-old *Zmpste24*^{+/-}*Lmna*^{+/-}, *Zmpste24*^{-/-}*Lmna*^{+/-} and *Zmpste24*^{-/-}*Lmna*^{-/-} mice. **b**, Body weight of 3-month-old *Zmpste24*^{+/-}*Lmna*^{+/-} (*n* = 9), *Zmpste24*^{-/-}*Lmna*^{+/-} (*n* = 10) and *Zmpste24*^{-/-}*Lmna*^{-/-} (*n* = 17) mice. **c**, Kaplan-Meier graph of *Zmpste24*^{+/-}*Lmna*^{+/-} (*n* = 7) and *Zmpste24*^{-/-}*Lmna*^{+/-} (*n* = 7) mice. The survival curve of *Zmpste24*^{-/-} mice is also shown (*n* = 45). **d**, Western blot analysis of lamins A and C in extracts from *Zmpste24*^{+/-}*Lmna*^{+/-}, *Zmpste24*^{-/-}*Lmna*^{+/-} and *Zmpste24*^{-/-}*Lmna*^{-/-} adult fibroblasts. **e**, SDS-PAGE of urine from *Zmpste24*^{+/-}*Lmna*^{+/-}, *Zmpste24*^{-/-}*Lmna*^{+/-} and *Zmpste24*^{-/-}*Lmna*^{-/-} male mice, showing recovery of production of major urinary proteins (MUPs). **f**, Senescence-associated β -galactosidase assay in kidney. Error bars represent s.e.m.; triple asterisk indicates *P* < 0.001.

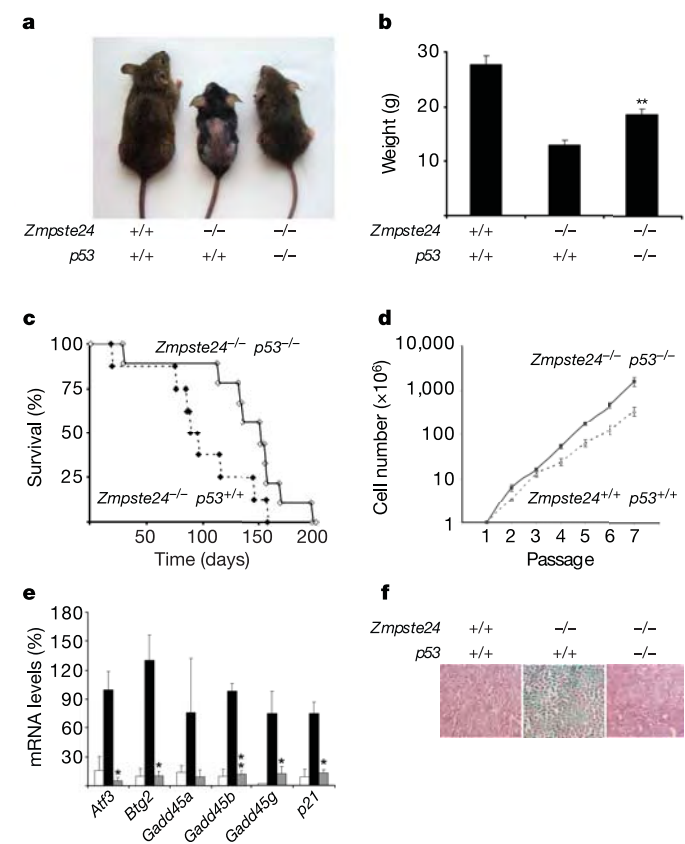


Figure 4 | Partial recovery of the *Zmpste24*^{-/-} phenotype in a p53-null background. **a**, Representative photograph of 3-month-old *Zmpste24*^{+/-}*p53*^{+/-}, *Zmpste24*^{-/-}*p53*^{+/-} and *Zmpste24*^{-/-}*p53*^{-/-} littermates. **b**, Total body weight of 3-month-old *Zmpste24*^{+/-}*p53*^{+/-} (*n* = 14), *Zmpste24*^{-/-}*p53*^{+/-} (*n* = 6) and *Zmpste24*^{-/-}*p53*^{-/-} (*n* = 9) mice. **c**, Kaplan-Meier survival graph of *Zmpste24*^{-/-}*p53*^{+/-} (*n* = 8) and *Zmpste24*^{-/-}*p53*^{-/-} (*n* = 9) mice. **d**, Proliferation assay of adult fibroblasts from control and *Zmpste24*^{-/-}*p53*^{-/-} mice. **e**, Transcriptional analysis by real-time quantitative PCR of p53 targets (and related genes) in grossly affected 3-month-old *Zmpste24*^{-/-}*p53*^{+/-} (black bars) (*n* = 3) mice and age-matched *Zmpste24*^{+/-}*p53*^{+/-} (white bars) (*n* = 3) and *Zmpste24*^{-/-}*p53*^{-/-} (grey bars) (*n* = 3) mice. mRNA levels are expressed as per cent values relative to a pool of *Zmpste24*^{-/-} liver RNAs. **f**, Senescence-associated β -galactosidase assay in kidney. Error bars represent s.e.m.; asterisk indicates *P* < 0.05; double asterisk indicates *P* < 0.01.

associated β -galactosidase staining was detected in kidneys from these double mutant mice. Lastly, quantitative PCR analysis of RNA levels of p53 targets in liver from severely affected *Zmpste24*^{-/-}*p53*^{+/+} and age-matched *Zmpste24*^{-/-}*p53*^{-/-} mice revealed decreased levels of *p21*, *Btg2*, *Gadd45a*, *Gadd45b*, *Gadd45g* and *Atf3* in the double knockout mice when compared with *Zmpste24*^{-/-}*p53*^{+/+} animals (Fig. 4). Taken together, these results demonstrate that the absence of p53 rescues some molecular alterations observed in *Zmpste24*^{-/-} mice. However, the fact that the multiple *Zmpste24*^{-/-} phenotype alterations were not rescued in a p53-null background to the same extent as in the case of the *Lmna*^{+/-} background indicates that other pathways contribute to the generation of the observed defects. In this regard, and consistent with previous findings in *Lmna*^{-/-} mice²⁴, it is of interest that retinoblastoma protein levels are reduced in *Zmpste24*^{-/-} mice (Supplementary Fig. 6). The microarray analysis also revealed changes in expression levels of a series of genes that were not described as p53 targets but that are of putative relevance for the observed phenotype in *Zmpste24*-null mice. This is the case for CD14 (upregulated 26-fold in the liver of *Zmpste24*^{-/-} mice), a receptor involved in apoptotic cell removal, the protein kinase Pim-3 (upregulated 13-fold), and the peroxisomal catalase (downregulated 6.1-fold) (Supplementary Table 1). Interestingly, levels of this antioxidant enzyme are reduced in fibroblasts of Hutchinson–Gilford progeria patients²⁵, its genetic ablation causes a progeric phenotype in *Caenorhabditis elegans*²⁶, and its overexpression extends murine lifespan²⁷. Therefore, the low levels of peroxisomal catalase in *Zmpste24*^{-/-} mice could contribute to the generation of some of the progeroid features observed in these mutant mice. Further studies will be required to identify the additional pathways that can cooperate with p53 in the generation of the multiple defects caused by accumulation of prelamin A in *Zmpste24*^{-/-} mutant mice. These studies may also provide additional information on the relevance, in both normal and pathological conditions, of the herein reported links between p53—the guardian of the genome—and lamin A—the guardian of the soma^{18,28}.

These phenotype rescue experiments support the conclusion that prelamin A accumulation in the nuclear envelope is the initial molecular event responsible for most or all phenotypic alterations, including progeroid features, observed in *Zmpste24*^{-/-} mice. These nuclear lamina alterations induce chromatin architecture changes that influence the regulation of gene expression through different signalling pathways, including those mediated by p53 activation. The finding of a p53 response linked to the activation of a cell senescence programme in *Zmpste24*^{-/-} progeroid mice could provide a connection of this model to recent work showing that mice producing a hyperactive p53 mutant protein or overexpressing a short isoform of p53 also exhibit accelerated ageing^{11,29}. Our results should also be consistent with the occurrence of a structural checkpoint that examines the integrity of the nuclear envelope and responds to putative damages in this structure by activating archetypal DNA damage responses such as those mediated by p53. Finally, the finding that the progeroid phenotypes in *Zmpste24*^{-/-} mice can be largely or partially rescued by lowering prelamin A levels or by targeting the p53 pathway activated by prelamin A accumulation, suggests that similar strategies might provide some therapeutic benefit to patients suffering from these devastating diseases.

METHODS

Animals. *Zmpste24*-deficient mice and *Lmna*-deficient mice were generated and genotyped as described¹⁴. *p53*-deficient mice³⁰ were provided by M. Serrano. Hepatic DNA damage was induced in 3-month-old mice by exposition to a single dose (10 Gy) of whole-body γ -irradiation. Animal experimentation was done in accordance with the guidelines of the Universidad de Oviedo.

Transcriptional profiling. Total RNA was isolated using an RNeasy kit (Qiagen). Double-stranded cDNA was synthesized using the SuperScript™ cDNA synthesis kit (Invitrogen). *In vitro* transcription was carried out with the Bioarray high yield RNA transcript labelling kit (Enzo Diagnostics). The biotin-labelled

cRNA was purified, fragmented and hybridized to murine genome U74Av2 GeneChips (Affymetrix).

Real-time quantitative PCR. Expression levels of selected genes (*Atf3*, *Btg2*, *Gadd45a*, *Gadd45b*, *Gadd45g* and *p21*) were analysed by using Applied Biosystems Tagman gene expression assays in an ABI7000 Sequence detection system (Applied Biosystems) following the manufacturer's instructions.

Western blotting and immunoprecipitation. Liver samples were homogenized in 50 mM Tris (pH 7.4), 150 mM NaCl, 1% NP-40, 50 mM NaF, 1 mM dithiothreitol, 2 mg ml⁻¹ pepstatin A, Complete inhibitor cocktail (Roche) and phosphatase inhibitor cocktails I and II (Sigma). p53 immunoprecipitation was performed with anti-p53 goat polyclonal antibody (FL-393-G, Santa Cruz). Immunoprecipitates were electrophoresed and transferred to nitrocellulose membranes. Blots were blocked with 3% non-fat dry milk, and incubated overnight at 4 °C with 1/5,000 anti-p53 rabbit polyclonal antibody CM1 (a gift of S. Laín), 1/100 anti-pRb (sc-102, Santa Cruz), 1/100 anti-p21 (sc-6246, Santa Cruz), 1/500 anti-laminA/C (MANLAC1, a gift from G. Morris), 1/100 anti-phosphohistone H2AX (05-636, Upstate), and 1/10,000 anti- β -actin (A5441, Sigma). Finally, blots were incubated with 1/1,000 goat anti-rabbit-HRP (Pierce) or 1/2,000 of anti-mouse-HRP (Amersham) in 1.5% non-fat milk, washed and developed with Femto chemiluminescent reagent (Pierce). To analyse levels of phosphorylated p53, membranes used to quantify total p53 were stripped and incubated with 1/1,000 rabbit polyclonal anti-phosphoserine-p53 (Ser 15)- or anti-phosphoserine-p53 (Ser 20)-specific antibodies (Cell Signaling).

Cell culture and proliferation assays. Fibroblasts were extracted from the ears of 12-week-old mice. Ears were sterilized with ethanol, washed with PBS and triturated with razor blades. Samples were then incubated with 600 μ l of 4 mg ml⁻¹ collagenase D (Roche) and 4 mg ml⁻¹ dispase II (Roche) in DMEM (Invitrogen) for 45 min at 37 °C and 5% CO₂. After filtering and washing, 6 ml of DMEM with 10% FBS (Invitrogen) and 1% antibiotic-antimycotic (Invitrogen) were added, and the mixture was incubated at 37 °C and 5% CO₂. 10⁶ cells were passed in a 10-cm plate every 3 days and cultured in 3% or 20% oxygen. To measure BrdU incorporation, cells were incubated for 4 h with 100 μ g μ l⁻¹ BrdU (Sigma), fixed with paraformaldehyde, and immunofluorescence was carried out with the B2531 antibody (Sigma).

Cell senescence and apoptosis assays. Senescence-associated β -galactosidase assays were carried out as described¹⁹. Apoptotic cells were measured with annexin-V FITC (BD Pharmingen) by flow cytometry. Immunohistochemical analyses were carried out in formalin-fixed tissues with anti-cleaved caspase-3 antibody (Cell Signaling).

Received 19 May; accepted 20 July 2005.

Published online 3 August 2005.

- Pendás, A. M. *et al.* Defective prelamin A processing and muscular and adipocyte alterations in *Zmpste24* metalloproteinase-deficient mice. *Nature Genet.* **31**, 94–99 (2002).
- Bergo, M. O. *et al.* *Zmpste24* deficiency in mice causes spontaneous bone fractures, muscle weakness, and a prelamin A processing defect. *Proc. Natl Acad. Sci. USA* **99**, 13049–13054 (2002).
- Corrigan, D. P. *et al.* Prelamin A endoproteolytic processing *in vitro* by recombinant *Zmpste24*. *Biochem. J.* **387**, 129–138 (2005).
- Sullivan, T. *et al.* Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J. Cell Biol.* **147**, 913–920 (1999).
- Mounkes, L. C., Kozlov, S., Hernandez, L., Sullivan, T. & Stewart, C. L. A progeroid syndrome in mice is caused by defects in A-type lamins. *Nature* **423**, 298–301 (2003).
- Agarwal, A. K., Fryns, J. P., Auchus, R. J. & Garg, A. Zinc metalloproteinase, ZMPSTE24, is mutated in mandibuloacral dysplasia. *Hum. Mol. Genet.* **12**, 1995–2001 (2003).
- De Sandre-Giovannoli, A. *et al.* Lamin A truncation in Hutchinson–Gilford progeria. *Science* **300**, 2055 (2003).
- Eriksson, M. *et al.* Recurrent de novo point mutations in lamin A cause Hutchinson–Gilford progeria syndrome. *Nature* **423**, 293–298 (2003).
- Chen, L. *et al.* LMNA mutations in atypical Werner's syndrome. *Lancet* **362**, 440–445 (2003).
- Navarro, C. L. *et al.* Loss of ZMPSTE24 (FACE-1) causes autosomal recessive restrictive dermopathy and accumulation of Lamin A precursors. *Hum. Mol. Genet.* **14**, 1503–1513 (2005).
- Tyner, S. D. *et al.* p53 mutant mice that display early ageing-associated phenotypes. *Nature* **415**, 45–53 (2002).
- Yu, J. *et al.* Identification and classification of p53-regulated genes. *Proc. Natl Acad. Sci. USA* **96**, 14517–14522 (1999).
- Qiu, W. *et al.* Hypermethylation of growth arrest DNA damage-inducible gene 45 beta promoter in human hepatocellular carcinoma. *Am. J. Pathol.* **165**, 1689–1699 (2004).
- Fei, P., Bernhard, E. J. & El-Deiry, W. S. Tissue-specific induction of p53 targets *in vivo*. *Cancer Res.* **62**, 7316–7327 (2002).

15. Wei, W., Hemmer, R. M. & Sedivy, J. M. Role of p14(ARF) in replicative and induced senescence of human fibroblasts. *Mol. Cell. Biol.* **21**, 6748–6757 (2001).
16. Bode, A. M. & Dong, Z. Post-translational modification of p53 in tumorigenesis. *Nature Rev. Cancer* **4**, 793–805 (2004).
17. Liu, B. *et al.* Genomic instability in laminopathy-based premature aging. *Nature Med.* **11**, 780–785 (2005).
18. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
19. Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
20. Campisi, J. Cancer and ageing: rival demons? *Nature Rev. Cancer* **3**, 339–349 (2003).
21. Serrano, M. & Blasco, M. A. Putting the stress on senescence. *Curr. Opin. Cell Biol.* **13**, 748–753 (2001).
22. Sharpless, N. E. & DePinho, R. A. Telomeres, stem cells, senescence, and cancer. *J. Clin. Invest.* **113**, 160–168 (2004).
23. Fong, L. G. *et al.* Heterozygosity for Lmna deficiency eliminates the progeria-like phenotypes in Zmpste24-deficient mice. *Proc. Natl Acad. Sci. USA* **101**, 18111–18116 (2004).
24. Johnson, B. R. *et al.* A-type lamins regulate retinoblastoma protein function by promoting subnuclear localization and preventing proteasomal degradation. *Proc. Natl Acad. Sci. USA* **101**, 9677–9682 (2004).
25. Yan, T., Li, S., Jiang, X. & Oberley, L. W. Altered levels of primary antioxidant enzymes in progeria skin fibroblasts. *Biochem. Biophys. Res. Commun.* **257**, 163–167 (1999).
26. Petriv, O. I. & Rachubinski, R. A. Lack of peroxisomal catalase causes a progeric phenotype in *Caenorhabditis elegans*. *J. Biol. Chem.* **279**, 19996–20001 (2004).
27. Schriener, S. E. *et al.* Extension of murine lifespan by overexpression of catalase targeted to mitochondria. *Science* **308**, 1909–1911 (2005).
28. Hutchison, C. J. & Worman, H. J. A-type lamins: guardians of the soma? *Nature Cell Biol.* **6**, 1062–1067 (2004).
29. Maier, B. *et al.* Modulation of mammalian life span by the short isoform of p53. *Genes Dev.* **18**, 306–319 (2004).
30. Jacks, T. *et al.* Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Astudillo for help with histopathological analysis; E. Fermiñán (CIC-Salamanca) for help with microarray experiments; X. S. Puente, A. Fueyo, J. Alvarez, P. Zuazua, G. Velasco, A. Bernad, S. Lain and M. Serrano for support and comments; T. Sánchez and L. Santos for help in animal care facilities; and M. Fernández, S. Alvarez and M. S. Pitiot for technical assistance. This work was supported by grants from Ministerio de Educación y Ciencia, Fundación “La Caixa”, the European Union, the Swedish Research Council, the Swedish Cancer Society and the Research Grant Council of Hong Kong. The Instituto Universitario de Oncología is supported by Obra Social Cajastur, and Red de Centros de Cancer Instituto Carlos III, Spain.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.L.-O. (clo@uniovi.es).

Insulin disrupts β -adrenergic signalling to protein kinase A in adipocytes

Jin Zhang^{1*†}, Christopher J. Hupfeld^{2*}, Susan S. Taylor^{3,4}, Jerrold M. Olefsky² & Roger Y. Tsien^{1,3,4}

Hormones mobilize intracellular second messengers and initiate signalling cascades involving protein kinases and phosphatases, which are often spatially compartmentalized by anchoring proteins to increase signalling specificity¹. These scaffold proteins may themselves be modulated by hormones^{2–4}. In adipocytes, stimulation of β -adrenergic receptors increases cyclic AMP levels and activates protein kinase A (PKA)⁵, which stimulates lipolysis by phosphorylating hormone-sensitive lipase and perilipin^{6–8}. Acute insulin treatment activates phosphodiesterase 3B, reduces cAMP levels and quenches β -adrenergic receptor signalling⁹. In contrast, chronic hyperinsulinaemic conditions (typical of type 2 diabetes) enhance β -adrenergic receptor-mediated cAMP production¹⁰. This amplification of cAMP signalling is paradoxical because it should enhance lipolysis, the opposite of the known short-term effect of hyperinsulinaemia. Here we show that in adipocytes, chronically high insulin levels inhibit β -adrenergic receptors (but not other cAMP-elevating stimuli) from activating PKA. We measured this using an improved fluorescent reporter and by phosphorylation of endogenous cAMP-response-element binding protein (CREB). Disruption of PKA scaffolding mimics the interference of insulin with β -adrenergic receptor signalling. Chronically high insulin levels may disrupt the close apposition of β -adrenergic receptors and PKA, identifying a new mechanism for crosstalk between heterologous signal transduction pathways.

In order to assess the effect of chronically high insulin levels on PKA activation (a signalling step immediately following cAMP production), we took advantage of genetically encoded A-kinase activity reporters (AKARs) for monitoring PKA activity in living cells. Such reporters serve as surrogate substrates for PKA and, when phosphorylated, generate a change in fluorescence resonance energy transfer (FRET) between two green fluorescent protein (GFP) mutants as the result of phosphorylation-induced intramolecular complex formation between a phosphoamino acid binding domain and the phosphorylated peptide¹¹. The previously described AKAR1 reporter has successfully monitored compartmentalized PKA activity. However, its poor sensitivity to cellular phosphatases impedes reversal of the FRET response, limiting the utility of AKAR1 for physiological studies.

To generate an improved AKAR reporter, we hypothesized that within the AKAR construct, tight binding of the phosphorylated peptide by the 14-3-3 module efficiently protects it from dephosphorylation¹². Thus, a phosphoamino acid binding domain that has reduced binding affinity for the phosphorylated peptide would be more suitable for constructing a reversible reporter. To this end, we chose the forkhead associated domain 1 (FHA1), a modular phosphothreonine binding domain with submicromolar binding affinity¹³, one order of magnitude weaker than that of 14-3-3 and

substrate peptides. Furthermore, peptide library screening has identified FHA1-binding peptides that favour specific amino acids around the phosphorylation site, particularly in the positions immediately C-terminal to the phospho-threonine (pT)¹³. We therefore modified the PKA substrate sequence to LRRATLVD by incorporating the near-optimal sequence for FHA1 binding at positions +1, +2 and +3 with respect to pT, while maintaining the PKA phosphorylation consensus motif at the –2, –3 positions (Fig. 1a).

We incorporated three linkers of different lengths between the substrate sequence and FHA1, and found that one construct provided a greater response and faster kinetics compared to the other two constructs (data not shown). This reporter is termed AKAR2, and when tested in HEK293 cells, it showed a reversible response. After the response induced by the β -adrenergic agonist isoproterenol reached a maximum, removal of the stimulant and addition of H89 (a relatively specific PKA inhibitor) led to a decrease in emission ratio, which returned to the basal level over 30–50 min (Fig. 1b). A second increase in emission ratio was generated by removal of H89 and addition of the adenylyl cyclase activator forskolin, showing that activation of AKAR2 was fully reversible. Pulsed photolytic release ('uncaging') of cAMP from a membrane-permeant precursor¹⁴ led to repeated cycles of increased and decreased emission ratios, presumably involving intracellular release then rapid degradation of cAMP followed by rapid changes in the activity balance between PKA and phosphatases (Fig. 1c).

To test the specificity of the reporter, we incubated AKAR2-expressing HEK293 cells with H89. As seen in Fig. 1d, this pretreatment prevented the response of AKAR2 to isoproterenol. The reporter itself was not compromised, as removal of H89 restored the ability of AKAR2 to respond to forskolin. Furthermore, co-expression of the specific PKA inhibitor PKI also abolished the response (Fig. 1e), confirming that the FRET response is PKA-specific. In addition, treatment of AKAR2-expressing HEK293 cells with phorbol dibutyrate (PDBu, a protein kinase C (PKC) activator) or thapsigargin (to stimulate calcium/calmodulin-dependent protein kinase II, CaMKII) did not alter the emission ratio (Fig. 1f, g). Thus, AKAR2 senses PKA but not PKC or CaMKII activation in living cells.

To confirm that changes in emission ratio result from phosphorylation of AKAR2 at the designated threonine residue, we changed the threonine to alanine in the PKA substrate sequence (LRRATLVD). Figure 1h shows that this single mutation, T391A, completely abolished the response to increases in cAMP. To compare the FRET change with the phosphorylation state of the reporter, immunoblotting analysis was performed using an antiphospho-PKA-substrate antibody before and after forskolin stimulation. AKAR2

¹Department of Pharmacology, ²Division of Endocrinology and Metabolism, ³Department of Chemistry and Biochemistry, and ⁴Howard Hughes Medical Institute, University of California at San Diego, La Jolla, California 92093, USA. [†]Present address: Department of Pharmacology and Molecular Sciences and Department of Neuroscience, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA.

*These authors contributed equally to this work.

migrated at the expected molecular weight, showing no proteolysis. Forskolin increased phosphorylation, correlating with the increase in FRET (Fig. 1i, lanes 1, 2). In contrast, no phosphorylation was detected for the variant T391A with or without forskolin stimulation (Fig. 1i, lanes 3, 4), indicating that the single threonine in the substrate sequence is the crucial phosphorylation site responsible for the emission ratio change.

As recently shown¹⁰, chronic insulin treatment leads to down-regulation of β -arrestin1 protein, thus impairing desensitization of β -adrenergic receptor (β -AR): G_s signalling. This results in a 25–63% increase in β -AR-mediated cAMP production compared with non-insulin-treated cells (Supplementary Fig. 1; see also ref. 10). As PKA is the major downstream target of cAMP, we measured the effect of insulin treatment on β -AR-mediated, cAMP-induced activation of PKA using AKAR2. In 3T3-L1 adipocytes microinjected with AKAR2 plasmid DNA, AKAR2 was expressed throughout the cell, and was excluded only from the lipid droplets (inset of Fig. 2b). Cells were pretreated with insulin for 8 h, and then acutely stimulated with varying concentrations of the β -adrenergic agonist isoproterenol. Insulin treatment (100 ng ml^{-1} for 8 h) did not alter basal AKAR2 phosphorylation (data not shown), but did delay PKA responses to isoproterenol stimulation, despite supersensitized cAMP production. As the isoproterenol dose decreased, these delays increased from 20 s ($10 \mu\text{M}$) to 150 s ($1 \mu\text{M}$) to 300 s ($0.1 \mu\text{M}$) (Fig. 2a, b).

β -AR signalling alters the expression of adipocyte-specific gene products such as leptin¹⁵, adiponectin¹⁶ and resistin¹⁷. Many transcriptional responses to PKA activation are mediated by PKA phosphorylation at Ser 133 of the cAMP response element binding protein (CREB)¹⁸, which is almost exclusively a nuclear protein. As shown in Fig. 2c, CREB was rapidly phosphorylated at Ser 133 upon

isoproterenol stimulation, and chronic insulin treatment inhibited this phosphorylation. The prolonged delay in CREB Ser 133 phosphorylation (relative to cytoplasmic AKAR2) is consistent with a 5–10 min delay in the response of the nuclear-localized AKAR2 (Supplementary Fig. 2). Thus, the antagonistic effect of chronic insulin treatment on β -AR activation of PKA is detectable by an endogenous effector protein (CREB), not just the exogenous fluorescent reporter.

Treatment of cells with the adenylyl cyclase activator forskolin or ultraviolet photolysis of caged cAMP raises intracellular cAMP and activates PKA independently of β -AR and G_s signalling. As shown in Fig. 3a, b, chronic hyperinsulinaemia did not delay AKAR2 phosphorylation induced by forskolin or by cAMP uncaging, suggesting that insulin treatment does not impair the total available pool of intracellular PKA. We conclude from these experiments that the insulin-induced delay in PKA activity is specific for a β -adrenergic-related pool of PKA.

Distinct pools of PKA can result from spatial compartmentalization of PKA activity by A-kinase anchoring proteins (AKAPs)¹, which bind to the regulatory subunit of PKA and direct the holoenzyme to discrete locations within the cell. To test whether AKAPs are involved in β -adrenergic-coupled PKA activation in adipocytes, we measured AKAR2 phosphorylation in the presence of Ht31, a synthetic peptide that blocks the interaction between AKAP and the regulatory subunit of PKA, thereby disrupting AKAP-mediated PKA targeting¹⁹. Ht31 and its inactive analogue Ht31p were microinjected as intact peptides. As shown in Fig. 4a, we found that microinjected Ht31 blocked the isoproterenol-stimulated FRET response, and slightly delayed but did not block forskolin-induced phosphorylation. On the other hand, the functionally inactive analogue Ht31p had no effect on the isoproterenol-stimulated FRET response (Supplementary Fig. 3).

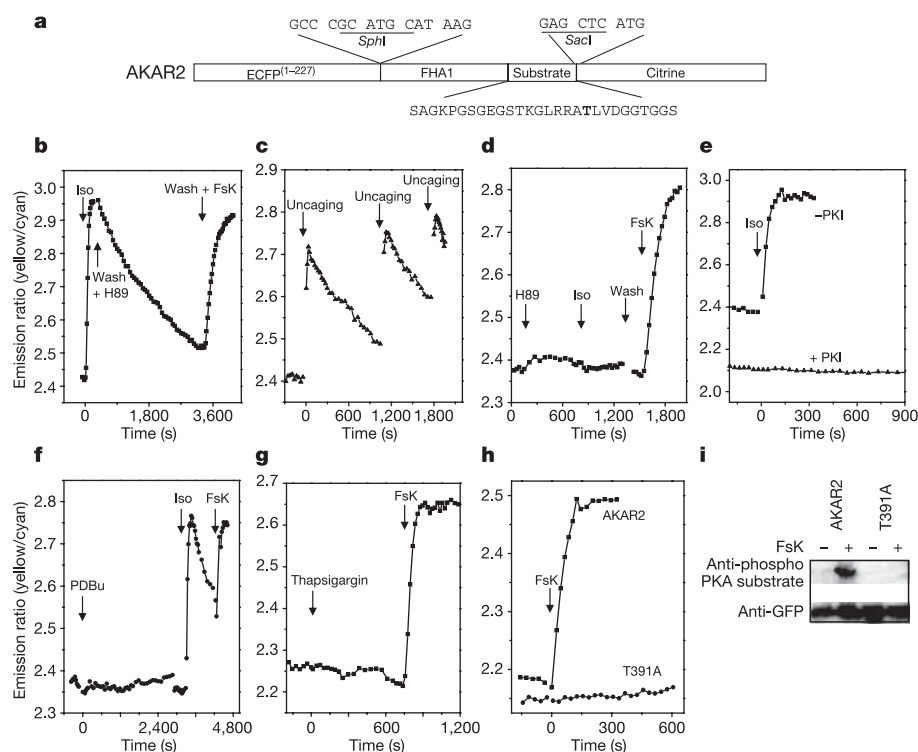


Figure 1 | Re-engineering of AKAR2. **a**, Domain structure of AKAR2. **b**, **c**, Representative FRET data in HEK293 cells in response to agonist stimulation (**b**) or cAMP uncaging (**c**). FSK, forskolin; Iso, isoproterenol. All time courses are representative of at least three experiments, total number of cells $n \geq 8$. **d**, **e**, Pretreatment with H89 (**d**) or co-expression of PKI (**e**, triangles, PKI; squares, control) blocks the response of AKAR2 to

isoproterenol. **f**, **g**, AKAR2 does not sense PKC (**f**) or CaMKII (**g**). **h**, FRET responses of the T391A mutant (circles) and wild-type AKAR2 (squares). **i**, Antiphospho-PKA-substrate (top) and anti-GFP (bottom) western blot analyses of wild-type and T391A AKAR2 from untreated and forskolin-treated cells.

To further examine the effect of chronic insulin treatment on PKA targeting, we measured the association between β_2 -AR and the RII β subunit of PKA, in the presence or absence of chronic insulin. We found that RII β and β_2 -AR were constitutively associated, and that this association was maintained or slightly increased after isoproterenol stimulation. Insulin pretreatment led to a pronounced decrease in β_2 -AR:RII β association after isoproterenol stimulation (Fig. 4b, c). These data suggest that AKAPs are involved in coupling a specific pool of PKA to β -AR, an association weakened by insulin pretreatment.

Chronic hyperinsulinaemia thus not only amplifies total cAMP production¹⁰, but more importantly, disrupts the linkage of β -ARs to a specific pool of PKA. The targeting of PKA to β -ARs by AKAPs^{20,21}, compartmentalized cAMP generation and PKA activation^{22,23} are well documented. Here we show that in adipocytes, compartmentalization of PKA with β -ARs is dynamically disrupted by insulin, providing a new mechanism for heterologous crosstalk. Previous examples of AKAP modulation²⁻⁴ were triggered by agents that directly increase cAMP levels and therefore represent homologous feedback. We propose that chronically high insulin levels weaken the close apposition of β -ARs and PKA (Supplementary Fig. 4) through modulation of AKAP scaffolding proteins, similar to the action of Ht31. Isoproterenol then seems to complete the dissociation between

β -AR and PKA in insulin-treated cells. In contrast, the association between β -AR and PKA in other systems is independent of or strengthened by agonist stimulation, as exemplified by AKAP79/150 (ref. 20) or gravin²⁴, respectively. We speculate that once chronic insulin treatment weakens the complex containing β -AR and PKA, PKA activation might phosphorylate one or more components and destroy the complex, thus terminating the β -AR/PKA association by very rapid negative feedback. The molecular components that constitute such signalling microdomains in adipocytes, and the mechanisms by which chronic hyperinsulinaemia affect their integrity, will be questions for future study. The improved AKARs presented here are reversible and targetable reporters allowing real-time imaging of PKA activity, and are valuable for analysing compartmentalized kinase activities.

METHODS

Gene construction. cDNA for enhanced cyan fluorescent protein (ECFP) and citrine²⁵ (an optimized version of YFP) were fused to forkhead associated domain 1 (FHA1) (Rad53p 22–162) created by polymerase chain reaction (PCR) using the Rad53-FHA1 template and primers containing the gene sequence for the linkers and phosphorylation substrate regions. The T391A mutation was incorporated by the QuickChange method (Stratagene). For expression in mammalian cells, the chimaeric proteins were subcloned into a modified pcDNA3 vector (Invitrogen) behind a Kozak sequence.

HEK293 cell culture and immunoblot analysis. HEK293 cells were plated onto sterilized glass coverslips in 3-cm dishes or 10-cm plates and grown to 50–90% confluency in DMEM medium supplemented with 10% FBS at 37 °C, 5% CO₂. Cells were then transfected using FuGENE-6 transfection reagent (Roche). For immunoblot analysis, cells expressing the reporters were stimulated with 50 μ M forskolin for 30 min at 25 °C. The cells were lysed with an ice-cold lysis buffer and the crude protein samples were concentrated, separated by SDS–polyacrylamide

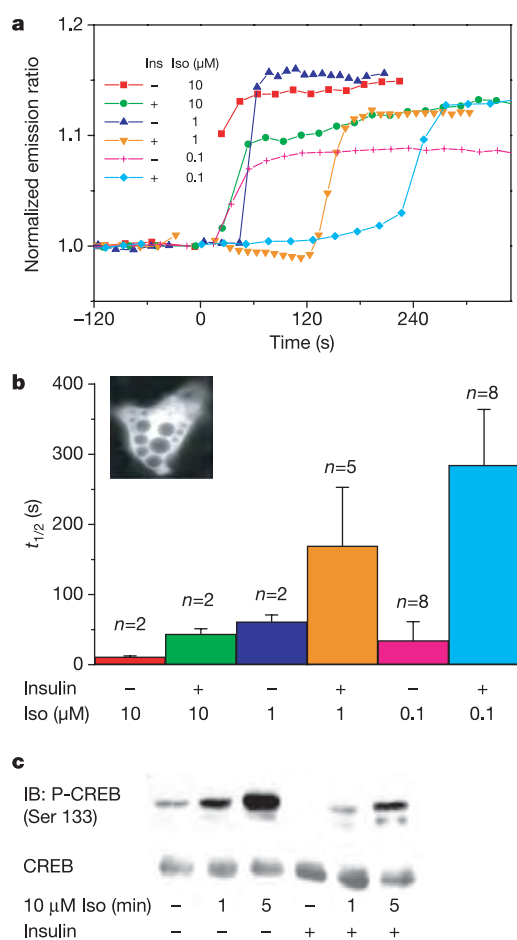


Figure 2 | Effect of insulin pretreatment on PKA activity. **a**, Representative data from cells stimulated with 10 μ M, 1 μ M or 0.1 μ M isoproterenol (Iso), with or without insulin (Ins) pretreatment. **b**, The mean time ($t_{1/2}$) for n cells to generate 50% of the maximal response with different doses of isoproterenol. Error bars indicate s.d. The inset shows a representative adipocyte expressing AKAR2. **c**, Effect of insulin treatment on CREB phosphorylation (P-CREB). Immunoblots (IB) are representative of two independent experiments.

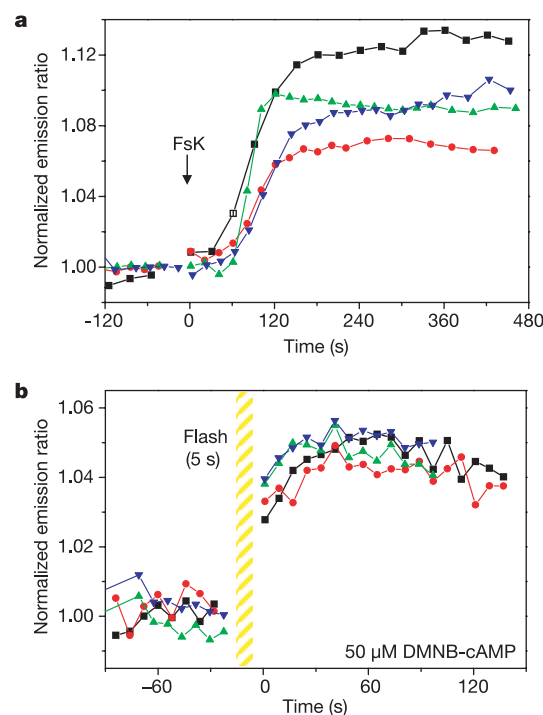


Figure 3 | The insulin-induced delay in PKA activity is specific for a β -adrenergic-coupled pool of PKA. **a**, Cells were stimulated acutely by the addition of forskolin. Insulin-treated cells in blue and green triangles ($t_{1/2} = 93 \pm 12$ s, mean $\pm$ s.d., $n = 8$); control cells in black squares and red circles ($t_{1/2} = 90 \pm 11$ s, $n = 7$). **b**, Cells were stimulated acutely by ultraviolet uncaging of DMNB-cAMP. Insulin-treated cells in blue and green triangles, representative data from three uncaging experiments ($n = 7$); control cells in black squares and red circles, representative data from three uncaging experiments ($n = 5$).

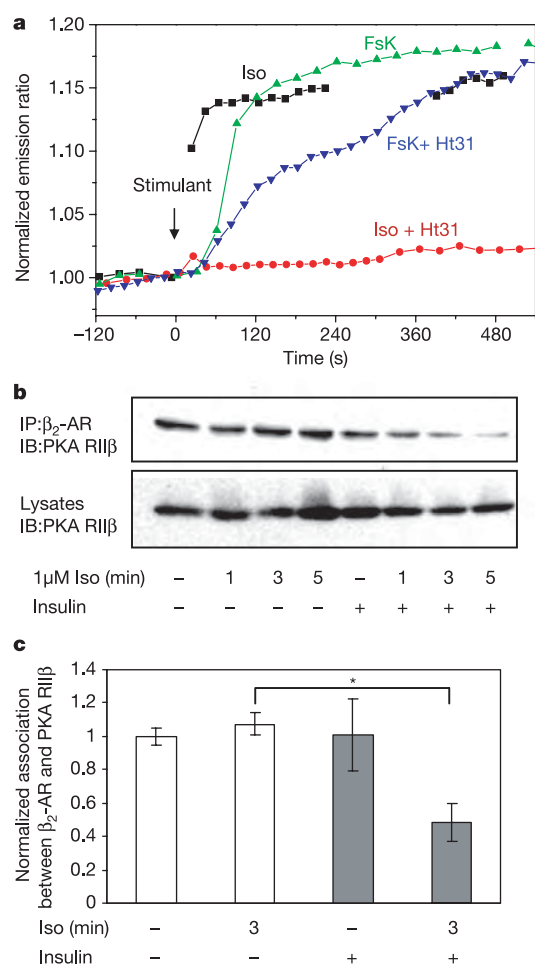


Figure 4 β -adrenergic-coupled pool of PKA. **a**, Representative data from cells microinjected with Ht31 and stimulated with 10 μ M isoproterenol ($n = 7$) or 50 μ M forskolin ($n = 6$). **b**, Effect of insulin pretreatment on the amount of PKA RII β co-immunoprecipitated with β_2 -AR. **c**, Quantification of PKA RII β co-immunoprecipitated with β_2 -AR, before (white) and after (grey) a 3-min treatment with isoproterenol. Data shown are mean $\pm$ s.e.m. from three experiments. Asterisk, $P = 0.045$, two-tailed t -test. Data are corrected for total amount of PKA RII β in lysates. Normalization is relative to the mean for cells without insulin or isoproterenol treatment.

gel electrophoresis and probed with an antiphospho-PKA-substrate antibody (New England Biolabs).

T3-L1 adipocyte culture, microinjection and immunoblot analysis. T3-L1 fibroblasts were grown to confluence at 37 °C, 10% CO₂, and differentiated into mature adipocytes as previously described²⁶. On day 7 post-differentiation, adipocytes were re-seeded to 3-cm dishes containing embedded glass coverslips, and microinjections were performed between days 10 and 14. AKAR2 expression plasmid was dissolved in microinjection buffer (5 mM NaH₂PO₄, 100 mM KCl, pH 7.2) to a final concentration of 0.1 mg ml⁻¹. Four hours before stimulation, 10 μ M Ht31 (DLIEEAASRIVDAVIEQVKAAGAY) was microinjected into nuclei along with the AKAR2 plasmid DNA. For analysing CREB phosphorylation, serum-starved T3-L1 adipocytes were treated, where indicated, with insulin (17 nM) for 8 h. After washes with PBS, cells were stimulated with isoproterenol (10 μ M) for 1 or 5 min. Proteins from cell lysates were probed on a western blot using a phospho-specific (Ser 133) anti-CREB antibody or an anti-CREB antibody.

Imaging. After transfection (24–72 h) or microinjection (4 h), the cells were washed twice with HBSS buffer, and maintained in buffer in the dark at 20–25 °C. Forskolin (Calbiochem) and isoproterenol (Aldrich) were then added as indicated. For inhibition studies of PKA activity, 10 μ M H89 was added to the cells before forskolin stimulation.

For uncaging of cAMP, cells were incubated with 50 μ M of a membrane-

permeant photolysable derivative, 4,5-dimethoxy-2-nitrobenzyl adenosine-3',5'-cyclic monophosphate (DMNB-cAMP), then exposed for 5 s to approximately 0.5 W cm⁻² of 340–370 nm illumination from the microscope's xenon lamp filtered through a 330WB80 filter. This ultraviolet light exposure was calculated to uncage >90% of the DMNB-cAMP within the cell.

Cells were imaged on a Zeiss Axiovert microscope with a $\times 40/1.3$ NA oil-immersion objective lens and a cooled CCD camera (Photometrics), controlled by Metafluor 2.75 software (Universal Imaging). Dual-emission ratio imaging used a 440DF30 excitation filter, a 455DRLP dichroic mirror and two emission filters (480DF30 for ECFP, 535DF25 for citrine) alternated by a filter changer (Lambda 10-2, Sutter Instruments). Fluorescence images were background-corrected. Exposure time was 100–1000 ms, and images were taken every 5–15 s (see Supplementary Methods).

Immunoprecipitation. T3-L1 adipocytes were treated with the indicated ligands, and whole-cell lysates were prepared. Lysate protein (500 μ g) was immunoprecipitated with 2 μ g of rabbit polyclonal anti- β_2 -AR antibody (Santa Cruz) overnight at 4 °C, and the resulting immune complexes were captured by incubating with Protein A beads for 2 h. After three washes with PBS, beads were dissolved in Laemmli buffer, boiled for 5 min, and proteins were separated by SDS gel electrophoresis and probed with an anti-RII β antibody (Biomol). Densitometry was performed using a Kodak ImageStation scanner and software.

Received 26 May; accepted 15 August 2005.

- Wong, W. & Scott, J. D. AKAP signalling complexes: focal points in space and time. *Nature Rev. Mol. Cell Biol.* **5**, 959–970 (2004).
- Salvador, L. M. et al. Neuronal microtubule-associated protein 2D is a dual A-kinase anchoring protein expressed in rat ovarian granulosa cells. *J. Biol. Chem.* **279**, 27621–27632 (2004).
- Carr, D. W., DeManno, D. A., Atwood, A., Hunzicker-Dunn, M. & Scott, J. D. Follicle-stimulating hormone regulation of A-kinase anchoring proteins in granulosa cells. *J. Biol. Chem.* **268**, 20729–20732 (1993).
- Feliciello, A., Rubin, C. S., Avvedimento, E. V. & Gottesman, M. E. Expression of a kinase anchor protein 121 is regulated by hormones in thyroid and testicular germ cells. *J. Biol. Chem.* **273**, 23361–23366 (1998).
- Taylor, S. S. et al. PKA: a portrait of protein kinase dynamics. *Biochim. Biophys. Acta* **1697**, 259–269 (2004).
- Londos, C., Honnor, R. C. & Dhillon, G. S. cAMP-dependent protein kinase and lipolysis in rat adipocytes. III. Multiple modes of insulin regulation of lipolysis and regulation of insulin responses by adenylate cyclase regulators. *J. Biol. Chem.* **260**, 15139–15145 (1985).
- Egan, J. J., Greenberg, A. S., Chang, M. K. & Londos, C. Control of endogenous phosphorylation of the major cAMP-dependent protein kinase substrate in adipocytes by insulin and beta-adrenergic stimulation. *J. Biol. Chem.* **265**, 18769–18775 (1990).
- Greenberg, A. S. et al. Perilipin, a major hormonally regulated adipocyte-specific phosphoprotein associated with the periphery of lipid storage droplets. *J. Biol. Chem.* **266**, 11341–11346 (1991).
- Elks, M. L. & Manganiello, V. C. Antilipolytic action of insulin: role of cAMP phosphodiesterase activation. *Endocrinology* **116**, 2119–2121 (1985).
- Hupfeld, C. J., Dalle, S. & Olefsky, J. M. β -Arrestin 1 down-regulation after insulin treatment is associated with supersensitization of β_2 adrenergic receptor G α s signalling in T3-L1 adipocytes. *Proc. Natl Acad. Sci. USA* **100**, 161–166 (2003).
- Zhang, J., Ma, Y., Taylor, S. S. & Tsien, R. Y. Genetically encoded reporters of protein kinase A activity reveal impact of substrate tethering. *Proc. Natl Acad. Sci. USA* **98**, 14997–15002 (2001).
- Muslin, A. J., Tanner, J. W., Allen, P. M. & Shaw, A. S. Interaction of 14-3-3 with signalling proteins is mediated by the recognition of phosphoserine. *Cell* **84**, 889–897 (1996).
- Durocher, D. et al. The molecular basis of FHA domain:phosphopeptide binding specificity and implications for phospho-dependent signalling mechanisms. *Mol. Cell* **6**, 1169–1182 (2000).
- Nerbonne, J. M., Richard, S., Nargeot, J. & Lester, H. A. New photoactivatable cyclic nucleotides produce intracellular jumps in cyclic AMP and cyclic GMP concentrations. *Nature* **310**, 74–76 (1984).
- Kosaki, A., Yamada, K. & Kuzuya, H. Reduced expression of the leptin gene (ob) by catecholamine through a Gs protein-coupled pathway in T3-L1 adipocytes. *Diabetes* **45**, 1744–1749 (1996).
- Fasshauer, M., Klein, J., Neumann, S., Eszlinger, M. & Paschke, R. Adiponectin gene expression is inhibited by β -adrenergic stimulation via protein kinase A in T3-L1 adipocytes. *FEBS Lett.* **507**, 142–146 (2001).
- Fasshauer, M., Klein, J., Neumann, S., Eszlinger, M. & Paschke, R. Isoproterenol inhibits resistin gene expression through a Gs-protein-coupled pathway in T3-L1 adipocytes. *FEBS Lett.* **500**, 60–63 (2001).
- Mayr, B. & Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nature Rev. Mol. Cell Biol.* **2**, 599–609 (2001).
- Scott, J. D. & Faux, M. C. Use of synthetic peptides in the dissection of protein-targeting interactions. *Methods Mol. Biol.* **88**, 161–185 (1998).

20. Fraser, I. D. *et al.* Assembly of an A kinase-anchoring protein- β_2 -adrenergic receptor complex facilitates receptor phosphorylation and signalling. *Curr. Biol.* **10**, 409–412 (2000).
21. Lin, F., Wang, H. & Malbon, C. C. Gravin-mediated formation of signalling complexes in β_2 -adrenergic receptor desensitization and resensitization. *J. Biol. Chem.* **275**, 19025–19034 (2000).
22. Jurevicius, J. & Fischmeister, R. cAMP compartmentation is responsible for a local activation of cardiac Ca^{2+} channels by β -adrenergic agonists. *Proc. Natl Acad. Sci. USA* **93**, 295–299 (1996).
23. Zaccolo, M. & Pozzan, T. Discrete microdomains with high concentration of cAMP in stimulated rat neonatal cardiac myocytes. *Science* **295**, 1711–1715 (2002).
24. Tao, J., Wang, H. Y. & Malbon, C. C. Protein kinase A regulates AKAP250 (gravin) scaffold binding to the β_2 -adrenergic receptor. *EMBO J.* **22**, 6419–6429 (2003).
25. Griesbeck, O., Baird, G. S., Campbell, R. E., Zacharias, D. A. & Tsien, R. Y. Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *J. Biol. Chem.* **276**, 29188–29194 (2001).
26. Dalle, S. *et al.* Insulin induces heterologous desensitization of G-protein-

coupled receptor and insulin-like growth factor I signalling by downregulating β -arrestin-1. *Mol. Cell. Biol.* **22**, 6272–6285 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.

Acknowledgements We thank Y. Ma for technical assistance and discussions and M. B. Yaffe for providing FHA cDNA. This work was supported by the Howard Hughes Medical Institute, the Alliance for Cell Signaling (R.Y.T.), the NIH (R.Y.T., C.J.H. and S.S.T) and the Johns Hopkins University School of Medicine and W. M. Keck Foundation (J.Z.).

Author Contributions J.Z. and C.J.H. performed all the experiments; all authors contributed to planning the experiments, interpreting the data and writing the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.Y.T. (rtsien@ucsd.edu).

LETTERS

The protein kinase A anchoring protein mAKAP coordinates two integrated cAMP effector pathways

Kimberly L. Dodge-Kafka^{1†}, Joseph Souhayer¹, Genevieve C. Pare², Jennifer J. Carlisle Michel¹, Lorene K. Langeberg¹, Michael S. Kapiloff² & John D. Scott¹

Cyclic adenosine 3', 5'-monophosphate (cAMP) is a ubiquitous mediator of intracellular signalling events. It acts principally through stimulation of cAMP-dependent protein kinases (PKAs)^{1,2} but also activates certain ion channels and guanine nucleotide exchange factors (Epacs)³. Metabolism of cAMP is catalysed by phosphodiesterases (PDEs)^{4,5}. Here we identify a cAMP-responsive signalling complex maintained by the muscle-specific A-kinase anchoring protein (mAKAP) that includes PKA, PDE4D3 and Epac1. These intermolecular interactions facilitate the dissemination of distinct cAMP signals through each effector protein. Anchored PKA stimulates PDE4D3 to reduce local cAMP concentrations, whereas an mAKAP-associated ERK5 kinase module suppresses PDE4D3. PDE4D3 also functions as an adaptor protein that recruits Epac1, an exchange factor for the small GTPase Rap1, to enable cAMP-dependent attenuation of ERK5. Pharmacological and molecular manipulations of the mAKAP complex show that anchored ERK5 can induce cardiomyocyte hypertrophy. Thus, two coupled cAMP-dependent feedback loops are coordinated within the context of the mAKAP complex, suggesting that local control of cAMP signalling by AKAP proteins is more intricate than previously appreciated.

Spatiotemporal control of cAMP flux requires the concerted action of adenylyl cyclases that synthesize cAMP, and phosphodiesterases that locally metabolize it into 5'-AMP (ref. 4). PDE binding proteins target distinct isozymes to specific subcellular sites^{5–7}. In cardiomyocytes, mAKAP assembles a negative feedback loop containing PDE4D3 and PKA at the perinuclear membrane⁸. PKA phosphorylation of Ser13 on PDE4D3 increases binding to mAKAP⁹, whereas phosphorylation (of PDE4D3 by PKA) of Ser54 enhances cAMP catabolism¹⁰. This configuration may generate local fluctuations in cAMP and pulses of compartmentalized PKA activity¹¹.

Here we used fluorescent reporters of PKA activity to test this model in cultured human and rat cells¹². AKAR2, a reversible A-kinase activity reporter³¹, is an improved version of AKAR1. This chimaeric protein consists of cyan fluorescent protein (CFP), a consensus PKA substrate sequence, a forkhead-associated (FHA) domain that binds phosphoamino acids, and the yellow fluorescent protein (YFP) citrine (Fig. 1a). PKA phosphorylation of the consensus site engages the FHA domain, permitting fluorescence resonance energy transfer (FRET) between the fluorescent moieties. Two modified AKAR2 reporters were generated: AKAR–PKA, which recruits PKA by means of a consensus anchoring sequence (Fig. 1b), and AKAR–PKA–PDE, which recruits PKA and tethers PDE4D3 through a binding site derived from mAKAP (Fig. 1c). HeLa cells expressing AKAP–PKA showed a rapid and sustained elevation of FRET on cAMP stimulation (Fig. 1b; red trace in Fig. 1d ($n = 10$); Supplementary Video 1). In contrast, cells expressing

AKAR–PKA–PDE showed a transient and less robust FRET response (Fig. 1c; green trace in Fig. 1d ($n = 13$); Supplementary Video 2). The PKA antagonist H89 blocked all cAMP-dependent FRET changes (Fig. 1d). PKA activity was unaffected by the PDE3 inhibitor milrinone (10 μ M). However, suppression of PKA activity was prevented when the same cells received the PDE4 inhibitor rolipram (1 μ M) (Fig. 1e, $n = 10$; Supplementary Video 3). Thus, recruitment of a PDE4 terminates activation of anchored PKA in the context of the AKAR–PKA–PDE reporter.

ERK kinases phosphorylate PDE4D3 on Ser579 to suppress phosphodiesterase activity¹³. We predicted that mAKAP might incorporate an ERK kinase module to counterbalance the anchored PDE4D3. Accordingly, sustained elevation of FRET on cAMP stimulation was observed in HeLa cells expressing AKAR–PKA–PDE plus constitutively active MEK, a kinase upstream of ERKs (black trace in Fig. 1f, $n = 11$). From this, we suggest that ERK activity suppresses anchored phosphodiesterase activity. Complementary studies were performed in cultured rat neonatal ventricular myocytes (RVNs). Treatment with 10% fetal calf serum activated ERK before immunoprecipitation of the mAKAP complex. The associated PDE activity was attenuated by $52 \pm 9\%$ ($n = 4$, $P < 0.0002$) compared to unstimulated controls (Fig. 1g, column 2). The MEK inhibitor PD98059 (20 μ M) blocked this effect (Fig. 1g, column 3, $P < 0.002$). Equivalent amounts of PDE4D3 were co-purified with the mAKAP anchoring protein under each experimental condition, suggesting that the reduction in phosphodiesterase activity is not a consequence of displacing the enzyme from the mAKAP complex (Fig. 1g, bottom panel). ERK activity was also measured in mAKAP-immunoprecipitated complexes (Fig. 1h, column 1), and an increase of 3.9 ± 0.7 -fold ($n = 3$, $P < 0.0007$) over the IgG control (immunoprecipitation with IgG) was observed (data not shown). PD98059 (2 μ M) blocked mAKAP-associated ERK activity (Fig. 1h, column 2, $P < 0.0001$). The PKA inhibitor PKI 5–24 peptide (50 nM) had no significant effect on ERK activity (Fig. 1h, column 3). Thus, mAKAP-associated ERK activity suppresses anchored PDE4D3 activity.

To determine which ERK family members might suppress mAKAP-associated PDE4D3, we performed western blot analysis on mAKAP immunoprecipitates from heart extracts. An 80-kDa protein that co-purified with mAKAP was detected with an antibody that recognizes all ERK family members and also with an ERK5-specific antibody (Fig. 1i, top and middle panels, lane 3). ERK5 was enriched in mAKAP immune complexes, but the 45-kDa ERK1/2 kinase was not (Fig. 1i, top panel, lane 1). The upstream kinase MEK5 (refs 14, 15) also co-precipitated with mAKAP (Fig. 1i, bottom panel). Immunofluorescence detection of mAKAP and ERK5 at the perinuclear membranes of hypertrophic RVNs confirmed the *in situ* assembly of this signalling complex¹⁶ (Supplementary Fig. S1).

¹Howard Hughes Medical Institute, Vollum Institute and ²Department of Pediatrics, Oregon Health and Sciences University, Portland, Oregon 97239, USA. [†]Present address: Calhoun Center for Cardiology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06030, USA.

We wanted to determine whether anchored PKA and ERK5 facilitate bidirectional regulation of the mAKAP-associated phosphodiesterase, hence controlling localized cAMP metabolism. Glutathione *S*-transferase (GST) pull-downs from heart extracts established the formation of a mAKAP–PDE4D3–ERK5 ternary complex (Supplementary Fig. S2). Cellular binding studies showed that PDE4D3 acts as an adaptor protein linking ERK5 to the mAKAP complex. This was confirmed when a mutant PDE4D3 form unable to bind ERK kinases (the KIM/FQF double mutant, see Methods¹⁷) could not recruit ERK5 to mAKAP (Supplementary Figs S3–S5). Thus, ERK5 is well positioned to suppress PDE4D3 activity and influence localized accumulation of cAMP.

cAMP modulates ERK signalling in a context-specific manner¹⁸. cAMP activates ERK in neuronally derived cell lines, but it inhibits ERK in certain fibroblasts and kidney cells^{19,20}. We evaluated cross-talk between mAKAP-anchored PKA and ERK5 in RNVs. When ERK was activated with 10% serum before immunoprecipitation of the mAKAP complex, mAKAP-associated ERK activity was increased 2.9 ± 0.6 -fold ($n = 3$, $P < 0.0003$) over controls (Fig. 2a, column 2). However, pretreatment with forskolin (20 μ M) to elevate intracellular cAMP resulted in an $89 \pm 8\%$ reduction in the serum-dependent activation of the mAKAP-associated ERK5 pool ($n = 3$, $P < 0.0004$; Fig. 2a, column 3). To our surprise, the PKA antagonist H89 (10 μ M) did not overcome the cAMP-dependent suppression of ERK5 activity (Fig. 2a, column 4). Similar results were obtained using two other well-characterized PKA inhibitors, KT5720 (1 μ M) and Rp-cAMPs (1 mM; Fig. 2a, columns 5, 6). Thus, the cAMP-dependent block of ERK5 activity was independent of PKA. Control experiments show that this effect was not a consequence of displacing ERK5 from the mAKAP complex (Fig. 2a, bottom). Therefore, another cAMP

effector within the mAKAP complex may function to suppress ERK5 activation.

One such candidate is Epac1 (exchange protein directly activated by cAMP), a cAMP-dependent guanine nucleotide exchange factor for the small G-protein Rap1 (refs 21, 22). Epac1 co-purified with mAKAP from heart extracts, but was not detected in an IgG control immunoprecipitation (Fig. 2b). Immunofluorescence techniques detected Epac1 at the perinuclear membranes of hypertrophic RNVs, where mAKAP is also found (Fig. 2c–e).

The cAMP analogue 8-CPT-2'-O-Me-cAMP was used to assess whether Epac1 is the cAMP effector that inhibits mAKAP-associated ERK5. This compound, also called 007, is a potent and selective activator of Epac proteins, with up to 300-fold selectivity over the R subunits of PKA²³. Treatment of RNVs with 007 alone resulted in the inhibition of ERK5 activity in mAKAP immune complexes (Fig. 2f, column 3, $n = 3$, $P < 0.0004$). KT5720 did not increase the effect of 007 (Fig. 2f, column 4), emphasizing that PKA does not have an additive effect on cAMP-dependent suppression of ERK5. Further biochemical experiments demonstrated that PDE4D3 recruits Epac1 to the mAKAP complex (Supplementary Figs S6–S8). This permits firm control of cAMP concentrations in the vicinity of Epac1, thereby regulating its guanine nucleotide exchange activity.

Epac1 is an effector protein of Rap1, a member of the Ras family of small G-proteins^{21,22}. Rap1 signalling is attenuated by the GTPase activating protein RapGAP²⁴. Adenovirus-mediated expression of constitutively active RapGAP in heart cells blocked Rap1 and, consequently, cAMP-mediated suppression of ERK5 activity (Fig. 2g, columns 1–3, $n = 4$, $P < 0.007$). Treatment with the PKA inhibitor KT5720 did not reverse this effect (Fig. 2g, column 4).

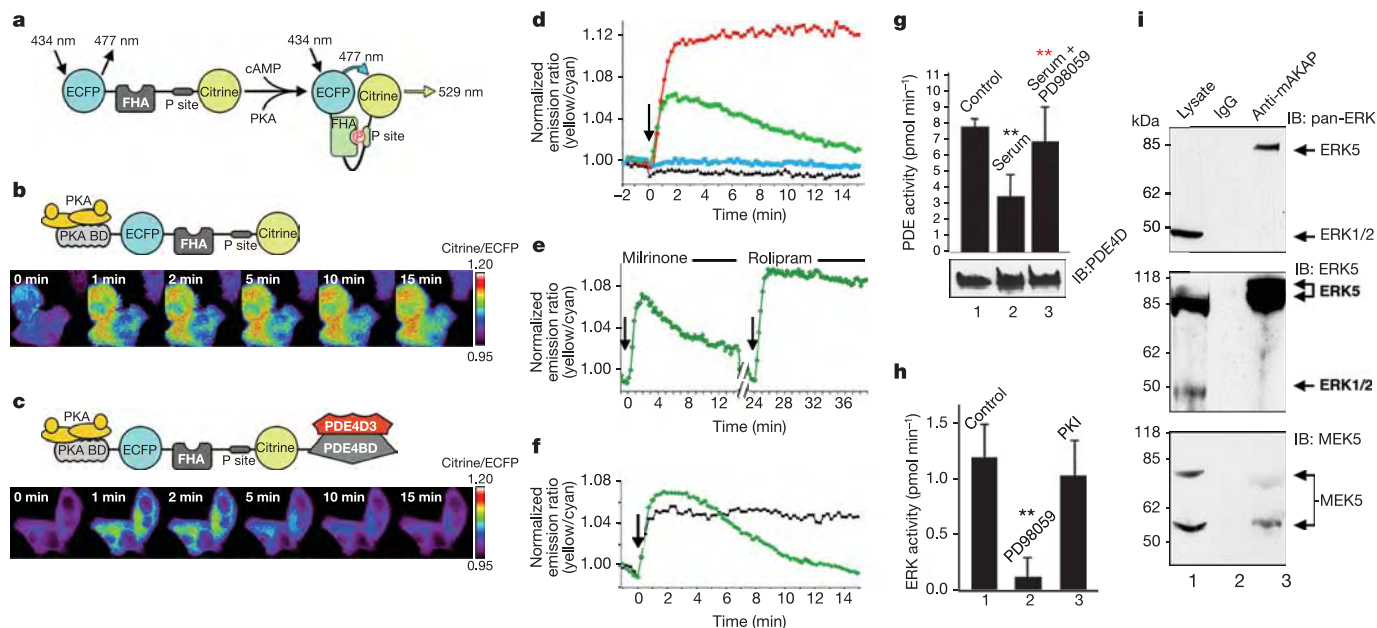


Figure 1 | Bidirectional control of the mAKAP-associated PDE4D3 activity. **a**, The modular composition and action of AKAR2. **b**, Schematic of AKAR2–PKA (top), and pseudo-coloured images of FRET changes in AKAR2–PKA-expressing HeLa cells stimulated with cAMP for 15 min (bottom). **c**, Schematic of AKAR2–PKA–PDE (top), and pseudo-coloured images of FRET changes in AKAR2–PKA–PDE-expressing HeLa cells stimulated with cAMP for 15 min (bottom). **d**, FRET measurements for AKAR–PKA (red, $n = 10$), AKAR–PKA–PDE (green, $n = 13$), AKAR–PKA + H89 (blue, $n = 10$) and AKAR–PKA–PDE + H89 (black, $n = 7$) for 15 min after cAMP stimulation with forskolin (arrow). **e**, FRET traces ($n = 10$) using the AKAR–PKA–PDE reporter after application of the PDE3 inhibitor milrinone (0–12 min) and the PDE4 inhibitor rolipram

(24–36 min). Stimulation with cAMP was at 0 and 24 min (arrows). **f**, FRET measurements from HeLa cells expressing the AKAR–PKA–PDE reporter in the presence (black, $n = 11$) or absence (green, $n = 7$) of a dominant active MEK5 for 15 min after cAMP stimulation with forskolin (arrow). **g**, Phosphodiesterase activity ($n = 4$) in mAKAP immune complexes isolated from RNVs. Treatment conditions are indicated. IB, immunoblot. **h**, ERK activity in mAKAP complexes immunoprecipitated from heart extracts ($n = 3$). Treatments with kinase inhibitors are indicated. Error bars in **g**, **h** show s.e.m. P values < 0.01 are indicated by asterisks (black, relative to control; red, relative to sample). **i**, Co-precipitated ERK was detected by immunoblot of mAKAP complexes from heart extracts using pan-ERK (top), ERK5-specific (middle) and MEK5-specific (bottom) antibodies.

Other control experiments demonstrated that adenoviral expression of β -galactosidase did not affect cAMP-dependent suppression of anchored ERK5 activity (Fig. 2h, $n = 3$, $P < 0.009$). Collectively, the data in Fig. 2 suggest that cAMP-dependent activation of Epac1 mobilizes a Rap1 signalling pathway, leading to inhibition of ERK5 in the mAKAP complex.

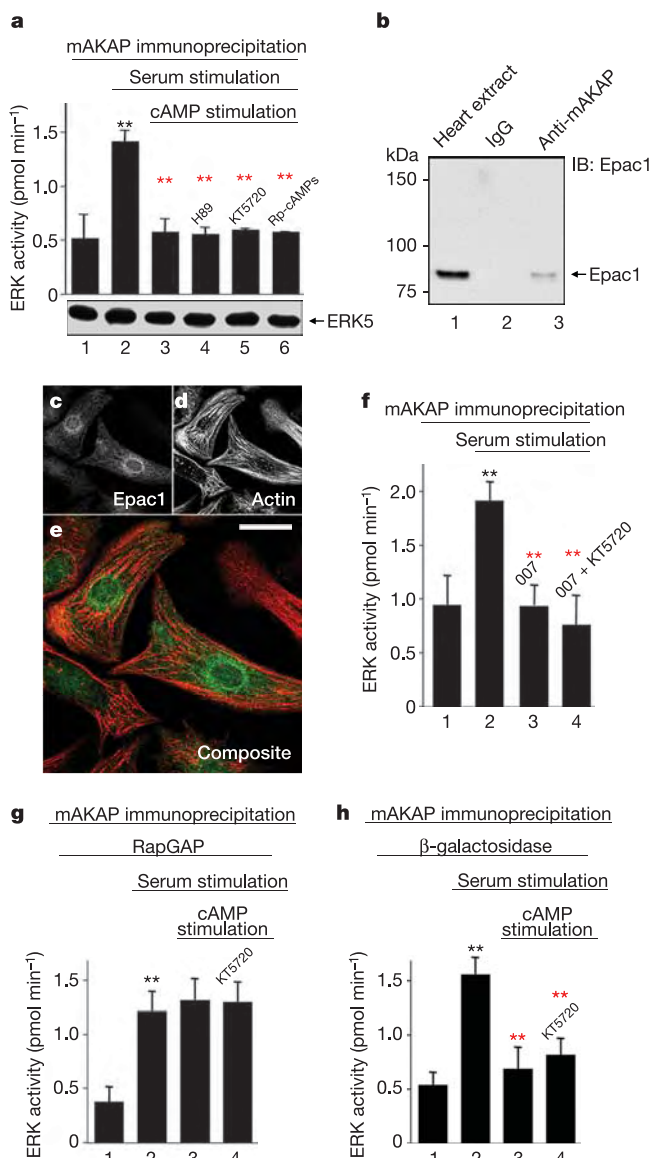


Figure 2 | Epac1 suppresses mAKAP-associated ERK5 activity. **a**, Serum-stimulated mAKAP-associated ERK5 activity ($n = 3$) in parallel cultures pretreated with forskolin to elevate cAMP (columns 2, 3) or in the presence of the PKA inhibitors H89, KT5720 or Rp-cAMPS (columns 4–6). Bottom panel shows the amount of ERK5 detected in each sample. **b**, Immunoblot detection of Epac1 in mAKAP immune complexes from rat heart extracts. **c–e**, Fluorescent staining of hypertrophic RNVs with an anti-Epac1 antibody (**c**) and Alexa 568 phalloidin to reveal the actin cytoskeleton (**d**). Composite image (**e**) shows the distribution of Epac1 (green) and actin (red). Scale bar, 20 μ m. **f**, The mAKAP complex was immunoprecipitated from cultured RNVs following serum stimulation to activate ERK. Before serum stimulation, parallel cultures were pretreated with either the Epac-selective activator 007 or KT5720 ($n = 3$). **g**, **h**, Serum-stimulated mAKAP-associated ERK activity in cells expressing constitutively active RapGAP (**g**, $n = 4$) or control β -galactosidase (**h**, $n = 3$). Stimulation of intracellular cAMP or treatment with the kinase inhibitor KT5720 is indicated. P values < 0.01 are indicated by asterisks (black, relative to control; red, relative to sample). Error bars in **f–h** show s.e.m.

Cytokines, including leukaemia inhibitory factor (LIF), activate ERK5 in RNVs to induce eccentric cardiac hypertrophy, an increase in cell size and length²⁵. LIF treatment for 24 h enhanced RNV cell size by 1.6-fold over controls (Fig. 3a–c, $n = 11$ independent experiments). This effect was blocked when ERK kinases were inhibited by PD98059 or if Epac1 was activated by 007, and was partially blocked on PDE4 inhibition with rolipram (Fig. 3c, columns 3–5). Plasmid-based RNA interference of mAKAP also suppressed LIF-mediated changes in RNV size compared with controls (Fig. 3d and Supplementary Fig. S9). The LIF response was rescued in RNVs

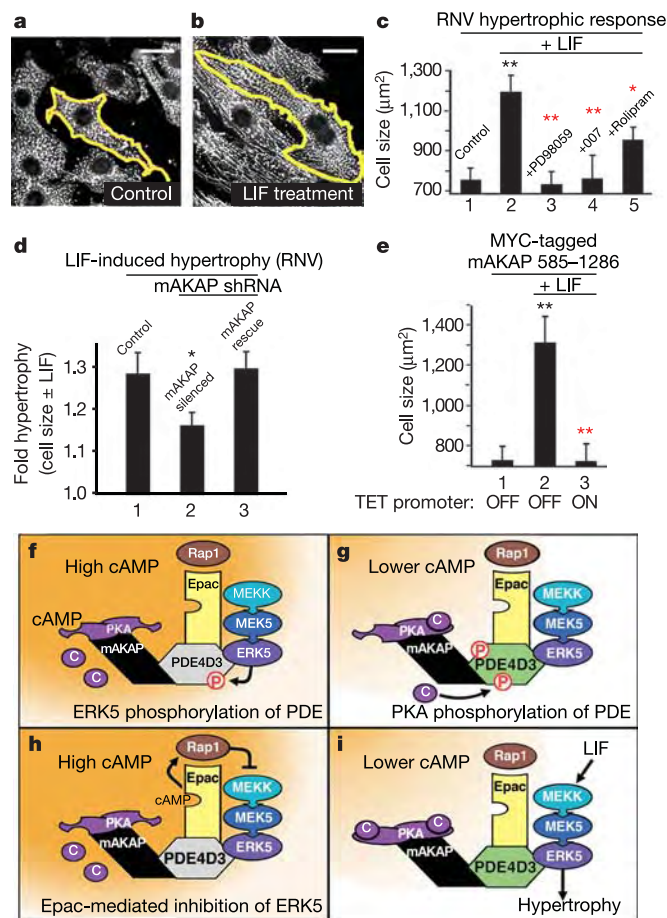


Figure 3 | The mAKAP complex facilitates cytokine-induced cardiac hypertrophy. **a–c**, Leukaemia inhibitory factor (LIF) induces changes in RNV size. The outline of a rat neonatal ventriculocyte from control (**a**) and LIF-treated (**b**) samples is presented. Scale bar, 20 μ m. **c**, Quantification of cell size in control RNVs (lane 1, $n = 11$) and LIF-treated RNVs (lane 2 ($n = 11$) and lanes 3–5). Pharmacological manipulation of ERK5 (PD98059, $n = 4$), Epac1 (007, $n = 4$) or PDE4 (rolipram, $n = 7$) activity is indicated. **d**, Expression of mAKAP was suppressed by RNA interference. Quantification of LIF-induced hypertrophy (cell size $\pm$ LIF) in control RNVs (lane 1, $n = 12$), mAKAP-silenced RNVs (lane 2, $n = 19$) and RNVs rescued with a variant of mAKAP resistant to the shRNA (lane 3, $n = 13$). **e**, Displacement of mAKAP from the nuclear membrane was achieved by overexpression of an mAKAP targeting domain fragment using the TET OFF inducible promoter. Quantification of cell size from control (lane 1, $n = 3$), from LIF-stimulated RNV controls (lane 2, $n = 4$), and from LIF-stimulated cells expressing the mAKAP 585–1286 fragment (lane 3, $n = 4$). P values are indicated by asterisks (black, relative to control; red, relative to sample): single asterisk, $P < 0.05$; two asterisks, $P < 0.01$. Error bars in **c–e** show s.e.m. **f–i**, Schematics highlighting the main findings of this study. **f**, ERK5 phosphorylation of PDE4D3 shuts down cAMP metabolism. **g**, PKA phosphorylation of PDE4D3 enhances cAMP metabolism. **h**, Activation of Epac mobilizes Rap1 to suppress ERK5 activation. **i**, Low cAMP represses the Epac-mediated block of ERK5, allowing cardiac hypertrophy.

expressing a variant of mA KAP resistant to the short hairpin (sh)RNA (Fig. 3d, column 3). Finally, displacement of mA KAP from the perinuclear membrane with an inducible competing mA KAP fragment (residues 585–1286) blocked LIF-induced RNV growth (Fig. 3e and Supplementary Fig. S10). Collectively, these data suggest that mA KAP organizes a cAMP-responsive network containing PDE4D3, Epac1 and ERK5 that modulates cardiomyocyte size. Similar results were observed when atrial natriuretic factor (ANF) expression was monitored as an independent index of hypertrophy (Supplementary Figs S11 and S12).

β -adrenergic agonists that elevate intracellular cAMP control the strength, duration and frequency of heart contraction. Cell imaging experiments in cardiomyocytes have revealed transient microdomains of cAMP where PKA becomes active^{26,27}. Our FRET data show that co-localization of PKA and a phosphodiesterase generate localized pulses of cAMP. A variety of cAMP-responsive events that have different durations and are responsive to different thresholds of cAMP can emanate from the same microdomain. This is particularly relevant for mA KAP signalling complexes, which contain three functionally distinct cAMP-dependent enzymes (PKA, PDE4D3 and Epac1). PKA responds to nanomolar concentrations of cAMP, and would become active early in a second messenger response. PDE4D3 (K_m 1–4 μ M) and Epac1 (K_d 4 μ M) would only become activated once cAMP concentrations reached micromolar levels⁴. Conversely, inactivation of PDE4D3 and Epac1 would precede PKA holoenzyme reformation as cAMP levels decline. Furthermore, Epac1 suppresses mA KAP-associated ERK5 activity in a PKA-independent manner. This is consistent with the idea that Epac proteins provide the cAMP-dependent link to ERK signalling events through Rap1^{22,23,28}.

In conclusion, mA KAP and PDE4D3 create an integrated and internally regulated signalling network. mA KAP anchors PKA and directs the cellular localization of the complex, while PDE4D3 serves as the adaptor protein for Epac1 and ERK5. The key regulatory enzyme is PDE4D3. ERK phosphorylation of PDE4D3 decreases the phosphodiesterase activity, thereby favouring local accumulation of cAMP and subsequent PKA activation (Fig. 3f) and Epac1 activation (Fig. 3h). Conversely, PKA phosphorylation of PDE4D3 increases its affinity for mA KAP⁹ and its V_{max} for cAMP¹⁰, decreasing the local cAMP concentration (Fig. 3g). This ultimately sustains ERK5 activity by repressing Epac1-mediated inhibition of ERK5 (Fig. 3i). This configuration creates a complex in which transduction enzymes from distinct pathways are spatially organized to exploit the temporal constraints of cAMP action.

METHODS

Antibodies. Primary antibodies used were monoclonal anti-pan-ERK (Transduction Labs), monoclonal anti-MEK5 (Transduction Labs), monoclonal anti-pan-PDE4D (ICOS), rabbit anti-ERK5 (StressGen), monoclonal anti-9E10-Myc (Upstate), rabbit anti-mA KAP (VO54), rabbit anti-Epac1 (Santa Cruz), mouse anti- α -actinin (Sigma) and rabbit anti-rat ANF (USB). Donkey HRP-conjugated (Jackson ImmunoResearch) and Alexa-conjugated (Molecular Probes) secondary antibodies were used.

AKAR–PKA and AKAR–PKA–PDE. Plasmids (pcDNA4) encoding AKAR–PKA or AKAR–PKA–PDE were derived from the original AKAR2 construct provided by R. Tsien. A DNA fragment encoding the consensus RII binding peptide from AKAPs (A-kinase anchoring protein-*in silico*: a 22-amino-acid peptide that is a high-affinity PKA-anchoring antagonist developed using a bioinformatic approach) following a Kozak sequence was generated by polymerase chain reaction (PCR). This fragment was ligated at the 5' end of AKAR2 to generate AKAR–PKA. The AKAR–PKA–PDE reporter was constructed by ligating a fragment encoding residues 1286–1831 of mA KAP to the 3' end of the AKAR–PKA coding region.

Cell culture. Preparation of primary RNVs and treatment to induce hypertrophy was as previously described¹⁶.

Live cell imaging. HeLa cells grown on glass coverslips were transfected with vectors encoding AKAR–PKA or AKAR–PKA–PDE using Lipofectamine 2000. After incubation at 37°C for 18–48 h, cells were washed twice with Hank's balanced salt solution, mounted in a Ludin chamber (Life Imaging Services) and

imaged using a Leica AS MDW workstation. Images were acquired using a Leica DM IRE2 microscope equipped with a CoolSNAP-HQ charge-coupled device camera (Roper Photometrics). This device was equipped with a Dual-view module (Optical Insights), D465/30 and HQ535/30 emission filters (Chroma) and a 505dcxr dichroic mirror (Chroma). CFP and YFP images were acquired simultaneously, aligned, background-corrected and analysed using FRET Applier software (Leica). Exposure time was from 250–1,000 ms, and images were taken every 15 s. Treatment with the PKA inhibitor H89 or isoform-specific PDE inhibitors preceded cAMP stimulation with forskolin by 10 min.

Expression constructs. Plasmids encoding wild-type PDE4D3 and mutant PDE4D3 (K455A/K456A/F597A/Q598A/F599A; this is the KIM/FQF double mutant) forms were from M. Houslay. pcDNA3 expression vectors encoding MYC-tagged ERK5 and Epac1, and a RAP-GAP adenovirus, were provided by P. Stork. For GST–PDE4D3 and GST–ERK5, complementary DNAs were subcloned into PGEX-4T2. shRNA vectors incorporated double-stranded hairpin oligonucleotides based on rat mA KAP messenger RNA sequence (NCBI GI:5070430, base pairs 7210–7228) under the direction of the U6 promoter: mA KAP shRNA (sense strand), 5'-GACGAACCTTCCTCCGAATTCAAGA GATTCGGAAGGAAGGTCGTCTTTT-3'; control shRNA (sense strand) 5'-GACGAACCCCTGTTCCGAATTCAAGAGATTCGGAACAGGGGTTCCG TCTTTT-3'. Bold base pairs in the control shRNA are different from the wild-type mA KAP shRNA. Rescue experiments were performed with a full-length MYC–His rat mA KAP lacking the shRNA recognition sequence that resides in the 3' untranslated region of the mA KAP message.

Enzyme assays. PDE activity was measured by the method in ref. 29, and PKA activity was measured as previously described⁸. ERK assays were performed on immunoprecipitates from heart extracts and RNVs. Immunoprecipitates were washed three times in HSE buffer before the addition of inhibitors, 25 μ l H₂O and 25 μ l ERK assay buffer (80 mM HEPES pH 7.4, 80 mM MgCl₂, 0.2 μ M ATP, 2 mM sodium orthovanadate, 20 mM NaF, 1 μ g tyrosine hydroxylase peptide (BioSource) and 10 μ Ci γ -³²ATP). After 30 min incubation at 30°C, the reaction mixture was spotted onto phosphocellulose strips and extensively washed in 75 mM phosphoric acid. Filters were washed in 95% ethanol, air-dried and counted by liquid scintillation.

RNVs were serum-starved for 16 h before the addition of either 20 μ M PD98059, the PKA inhibitors H 89 (10 μ M), KT5720 (1 μ M) or Rp-cAMPs (1 mM), for 1 h. Forskolin (20 μ M) and 3-isobutyl-1-methylxanthine (IBMX, 75 μ M) were then added for 20 min. For 8-CPT-2'-O-Me-cAMP treatment (50 μ M, the Epac-selective cAMP analogue 007), cells were not exposed to forskolin. Cells were stimulated with 10% fetal calf serum for 10 min, lysed in 1 ml HSE buffer containing 50 mM NaF, 1 mM sodium orthovanadate, 0.1 μ M okadaic acid and 0.1 μ M cyclosporine, and immunoprecipitations were performed as described above.

H9C2 heart cells were incubated with adenoviral constructs of Rap1GAP1 or β -galactosidase at a multiplicity of infection (MOI) of 50 for 48 h. The cells were treated as described above before the addition of forskolin/IBMX and before serum stimulation.

Immunocytochemistry and microscopy. Fixation and staining of RNVs were as previously described⁸. To determine the effect of shRNA expression on cellular hypertrophy, myocytes were co-transfected with pTRE-U6 mA KAP shRNA or control shRNA plasmid, pEGFPN3 green fluorescent protein expression plasmid (Clontech) and either pBluescript carrier DNA or pcDNA3.1 MYC–His mA KAP expression vector for rescue experiments. Cells were stained with primary and Alexa-conjugated specific secondary antibodies. Anti- α -actinin stained sarcomeric Z-disks to distinguish myocytes from contaminating fibroblasts. Images were acquired using a Leica DMRA fluorescent microscope. Cell size was measured using the method in ref. 30. Data are presented as the average mean cell area $\pm$ s.e.m. *P*-values were determined using a two-tailed Student's *t*-test.

Received 5 May; accepted 27 June 2005.

- Walsh, D. A., Perkins, J. P. & Krebs, E. G. An adenosine 3',5'-monophosphate-dependent protein kinase from rabbit skeletal muscle. *J. Biol. Chem.* **243**, 3763–3765 (1968).
- Su, Y. *et al.* Regulatory subunit of protein kinase A: structure of deletion mutant with cAMP binding proteins. *Science* **269**, 807–813 (1995).
- Bos, J. L. Epac: a new cAMP target and new avenues in cAMP research. *Nature Rev. Mol. Cell Biol.* **4**, 733–738 (2003).
- Beavo, J. A. & Brunton, L. L. Cyclic nucleotide research—still expanding after half a century. *Nature Rev. Mol. Cell Biol.* **3**, 710–718 (2002).
- Houslay, M. D. & Adams, D. R. PDE4 cAMP phosphodiesterases: modular enzymes that orchestrate signalling cross-talk, desensitization and compartmentalization. *Biochem. J.* **370**, 1–18 (2003).
- Perry, S. J. *et al.* Targeting of cyclic AMP degradation to β 2-adrenergic receptors by β -arrestins. *Science* **298**, 834–836 (2002).

7. Verde, I. *et al.* Myomegalin is a novel protein of the Golgi/centrosome that interacts with a cyclic nucleotide phosphodiesterase. *J. Biol. Chem.* **276**, 11189–11198 (2001).
8. Dodge, K. L. *et al.* mAKAP assembles a protein kinase A/PDE4 phosphodiesterase cAMP signalling module. *EMBO J.* **20**, 1921–1930 (2001).
9. Carlisle Michel, J. J. *et al.* PKA phosphorylation of PDE4D3 facilitates recruitment of the mAKAP signalling complex. *Biochem. J.* **381**, 587–592 (2004).
10. Sette, C. & Conti, M. Phosphorylation and activation of a cAMP-specific phosphodiesterase by the cAMP-dependent protein kinase. *J. Biol. Chem.* **271**, 16526–16534 (1996).
11. Wong, W. & Scott, J. D. AKAP signalling complexes: Focal points in space and time. *Nature Rev. Mol. Cell Biol.* **5**, 959–971 (2004).
12. Zhang, J., Ma, Y., Taylor, S. S. & Tsien, R. Y. Genetically encoded reporters of protein kinase A activity reveal impact of substrate tethering. *Proc. Natl Acad. Sci. USA* **98**, 14997–15002 (2001).
13. Hoffmann, R., Baillie, G. S., MacKenzie, S. J., Yarwood, S. J. & Houslay, M. D. The MAP kinase ERK2 inhibits the cyclic AMP-specific phosphodiesterase HSPDE4D3 by phosphorylating it at Ser579. *EMBO J.* **18**, 893–903 (1999).
14. Zhou, G., Bao, Z. Q. & Dixon, J. E. Components of a new human protein kinase signal transduction pathway. *J. Biol. Chem.* **270**, 12665–12669 (1995).
15. English, J. M., Vanderbilt, C. A., Xu, S., Marcus, S. & Cobb, M. H. Isolation of MEK5 and differential expression of alternatively spliced forms. *J. Biol. Chem.* **270**, 28897–28902 (1995).
16. Kapiloff, M. S., Schillace, R. V., Westphal, A. M. & Scott, J. D. mAKAP: an A-kinase anchoring protein targeted to the nuclear membrane of differentiated myocytes. *J. Cell Sci.* **112**, 2725–2736 (1999).
17. MacKenzie, S. J., Baillie, G. S., McPhee, I., Bolger, G. B. & Houslay, M. D. ERK2 mitogen-activated protein kinase binding, phosphorylation, and regulation of the PDE4D cAMP-specific phosphodiesterases. The involvement of COOH-terminal docking sites and NH₂-terminal UCR regions. *J. Biol. Chem.* **275**, 16609–16617 (2000).
18. Pearson, G. W. & Cobb, M. H. Cell condition-dependent regulation of ERK5 by cAMP. *J. Biol. Chem.* **277**, 48094–48098 (2002).
19. Cook, S. J. & McCormick, F. Inhibition by cAMP of Ras-dependent activation of Raf. *Science* **262**, 1069–1072 (1993).
20. Vaillancourt, R. R., Gardner, A. M. & Johnson, G. L. B-Raf-dependent regulation of the MEK-1/mitogen-activated protein kinase pathway in PC12 cells and regulation by cyclic AMP. *Mol. Cell Biol.* **14**, 6522–6530 (1994).
21. de Rooij, J. *et al.* Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature* **396**, 474–477 (1998).
22. Kawasaki, H. *et al.* A family of cAMP-binding proteins that directly activate Rap1. *Science* **282**, 2275–2279 (1998).
23. Enserink, J. M. *et al.* A novel Epac-specific cAMP analogue demonstrates independent regulation of Rap1 and ERK. *Nature Cell Biol.* **4**, 901–906 (2002).
24. Rubinfeld, B. *et al.* Molecular cloning of a GTPase activating protein specific for the Krev-1 protein p21rap1. *Cell* **65**, 1033–1042 (1991).
25. Nicol, R. L. *et al.* Activated MEK5 induces serial assembly of sarcomeres and eccentric cardiac hypertrophy. *EMBO J.* **20**, 2757–2767 (2001).
26. Zaccolo, M. & Pozzan, T. Discrete microdomains with high concentration of cAMP in stimulated rat neonatal cardiac myocytes. *Science* **295**, 1711–1715 (2002).
27. Mongillo, M. *et al.* Fluorescence resonance energy transfer-based analysis of cAMP dynamics in live neonatal rat cardiac myocytes reveals distinct functions of compartmentalized phosphodiesterases. *Circ. Res.* **95**, 67–75 (2004).
28. Larocche-Joubert, N., Marsy, S., Michelet, S., Imbert-Teboul, M. & Doucet, A. Protein kinase A-independent activation of ERK and H,K-ATPase by cAMP in native kidney cells: role of Epac I. *J. Biol. Chem.* **277**, 18598–18604 (2002).
29. Beavo, J. A., Bechtel, P. J. & Krebs, E. G. Preparation of homogeneous cyclic AMP-dependent protein kinase(s) and its subunits from rabbit skeletal muscle. *Methods Enzymol.* **38**, 299–308 (1974).
30. Kodama, H. *et al.* Significance of ERK cascade compared with JAK/STAT and PI3-K pathway in gp130-mediated cardiac hypertrophy. *Am. J. Physiol. Heart Circ. Physiol.* **279**, H1635–H1644 (2000).
31. Zhang, J., Hupfeld, C. J., Taylor, S. S., Olefsky, J. M. & Tsien, R. Y. Insulin disrupts β -adrenergic signalling to protein kinase A in adipocytes. *Nature* doi:10.1038/nature04140 (this issue).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institutes of Health (to J.D.S. and M.S.K.) and the American Heart Association (to K.L.D.-K.). The authors wish to thank N. Mayer, D. Bleckinger and R. Mouton for technical assistance, and R. Tsien, M. Houslay, J. E. Dixon, J. L. Bos and P. Stork for reagents.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.D.S. (scott@ohsu.edu).

Natural-like function in artificial WW domains

William P. Russ¹, Drew M. Lowery², Prashant Mishra¹, Michael B. Yaffe² & Rama Ranganathan¹

Protein sequences evolve through random mutagenesis with selection for optimal fitness¹. Cooperative folding into a stable tertiary structure is one aspect of fitness, but evolutionary selection ultimately operates on function, not on structure. In the accompanying paper², we proposed a model for the evolutionary constraint on a small protein interaction module (the WW domain) through application of the SCA, a statistical analysis of multiple sequence alignments^{3,4}. Construction of artificial protein sequences directed only by the SCA showed that the information extracted by this analysis is sufficient to engineer the WW fold at atomic resolution. Here, we demonstrate that these artificial WW sequences function like their natural counterparts, showing class-specific recognition of proline-containing target peptides^{5–8}. Consistent with SCA predictions, a distributed network of residues mediates functional specificity in WW domains. The ability to recapitulate natural-like function in designed sequences shows that a relatively small quantity of sequence information is sufficient to specify the global energetics of amino acid interactions.

The basic premise of the SCA is that in accord with the cooperative nature of amino acid interactions in determining protein stability

and function, the evolutionary constraint on proteins should be a distributed (rather than intrinsic) property of amino acid positions. That is, the conservation of amino acids at one site should be the result of constraints on that site taken independently, and of the constraints arising through energetic interactions with other positions. For example, consider the SCA for an alignment of 292 members of the WW domain family (Fig. 1). This small protein interaction module adopts a curved three-stranded β -sheet structure with a binding site for proline-containing peptides formed on the concave surface of the sheet (Fig. 1a). The binding surface includes an X-Pro binding site (positions 19 and 30, shown in blue as CPK representation; Fig. 1a) that recognizes the canonical proline in target peptides, and a specificity site formed by residues in $\beta 2$ and the $\beta 2$ – $\beta 3$ loop (positions 21, 23 and 26, shown in yellow as CPK representation; Fig. 1a)⁹. WW domains are classified into four groups based on target peptide sequence motifs: group I, PPxY⁶; group II, PPLP⁷; group III, PPR⁵; and group IV, pS/pT-P⁸, where x stands for any amino acid. The SCA output for the WW family is a matrix of coupling values organized such that columns represent positions in the WW alignment, and rows represent sites where perturbations are

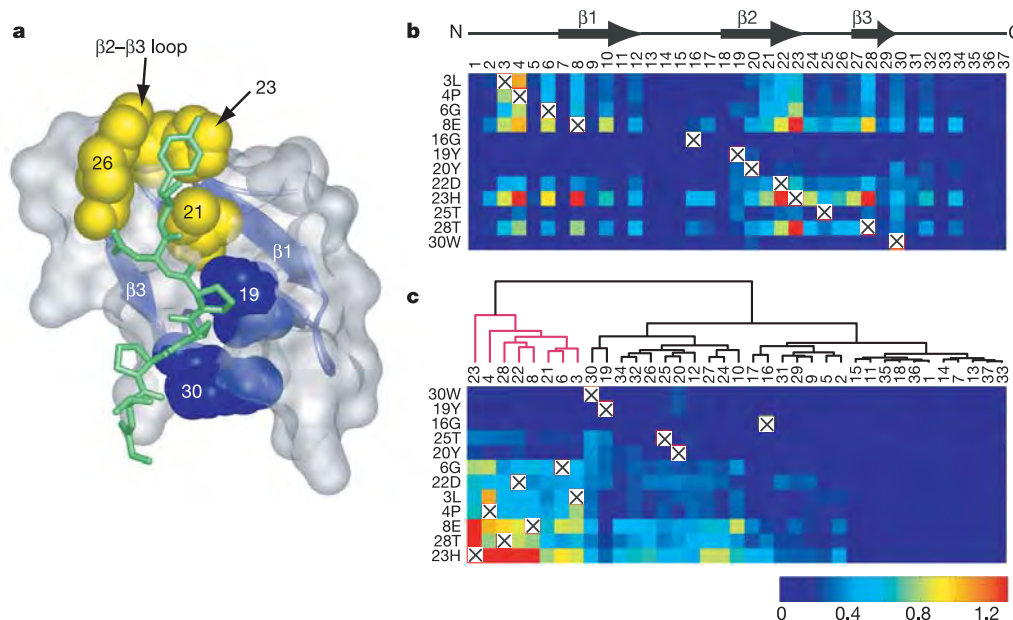


Figure 1 | Statistical coupling analysis for the WW domain, a proline-binding protein interaction module. **a**, The peptide-binding surface of the WW domain contains two structurally defined pockets: the X-Pro binding site (in blue) and a specificity site (in yellow). Shown is the Nedd4.3 WW domain (Protein Data Bank 1I5H) bound to a group I peptide (in green)³⁰. **b**, A matrix of coevolution scores between all WW positions (columns) and

12 moderately conserved positions (rows) in an alignment of 292 WW domains. The colour scale ranges from blue (no coevolution) to red (maximal coevolution). Checked boxes mark the trivial self-correlation of sites in the SCA method. **c**, Hierarchical clustering shows that a single group of eight positions (marked in red) share a pattern of strong mutual coevolution.

¹Howard Hughes Medical Institute and Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9050, USA. ²Center for Cancer Research, Department of Biology and Division of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

introduced to interrogate evolutionary coupling⁴ (Fig. 1b). Thus, each pixel shows the coevolution score for one pair of WW sites. Hierarchical clustering of this matrix reveals a remarkably simple global organization of conserved evolutionary interactions between amino acids. WW positions fall into two main clusters, one that shows minimal coupling to other sites, and one that comprises eight positions (marked in red) related by strong mutual coevolution (Fig. 1c).

Is the amino acid composition at sites plus the information in the SCA matrix all the sequence information required to specify the WW domain? The accompanying paper provides the first phase of this experimental test by showing that artificial sequences designed using only these parameters adopt a stable WW-like tertiary structure². However, a key further test is the sufficiency of this information for specifying function. If the SCA matrix captures the fitness constraints on the WW family, then the artificial sequences should show class-specific recognition of proline-containing sequences and binding affinities like those of natural WW domains.

We developed an oriented peptide library binding assay for measuring WW domain specificity, and studied a set of natural and artificial sequences. Four biotinylated degenerate peptide libraries were constructed, each oriented around one group-specific WW recognition motif, and binding was detected using an enzyme-linked immunosorbent assay (ELISA) (Fig. 2a, b; see also Supplementary Fig. 1a). For example, the group-I-oriented peptide library was biotin-Z-GMAxxxPPxYxxxAKKK, where Z is 6-amino-hexanoic acid and x stands for any amino acid except cysteine (theoretical degeneracy of 8.9×10^8 sequences). A fifth proline-oriented library was also made as a control for nonspecific binding. For a group I domain (Nedd4.3, or N39 by the numbering system in the accompanying paper) the peptide library assay confirms specific interaction with the PPxY-containing sequences¹⁰ (Fig. 2a). Interestingly, CC43, an artificial WW domain created through SCA-based protein design², also specifically interacts with the PPxY peptide library (Fig. 2b), suggesting that CC43 functions as a group I WW domain. To confirm these results independently, we carried out phage-display and fluorescence-based quantitative binding assays for N39 and CC43. Consistent with the peptide library screens, both N39 and CC43 selectively isolate PPxY-containing sequences from a random 12-mer phage display library (Fig. 2c, d). In addition, both of these domains show similar binding affinities to a model group I peptide (N39, dissociation constant (K_d) = $11.2 \pm 1.2 \mu\text{M}$; CC43, K_d = $1.7 \pm 0.1 \mu\text{M}$; mean $\pm$ s.d.; Fig. 2e). Taken together, these data validate the oriented peptide library assay for classification of WW domain specificity, and demonstrate that one artificial WW domain, CC43, displays group I-specific binding.

Figure 3 summarizes the results of peptide library screening for 27 randomly chosen natural and ten natively folded artificial WW domains². The Pin1 WW domain was added to the list of natural domains tested because it is the best-characterized member of the group IV specificity class¹¹. The data are shown in clustered matrix format, with each row showing the binding for one domain to all five oriented peptide libraries normalized between the minimum (white) and the maximum (black) observed signal to allow comparison of specificity profiles. Group I WW domains fell into two distinct subclusters (I and Im (marginal group I)) that significantly differed in the degree of specificity for the PPxY motif library (Supplementary Fig. 1). Group III domains bound to the PPR-oriented peptide library as expected but also showed binding to the PPxY library and weak binding to PPLP and proline-alone libraries. The binding to the PPxY library may simply reflect the fact that this library contains a non-trivial fraction of PPRY sequences. Nevertheless, the overall profiles show that group III domains display far more relaxed specificity than group I domains (Supplementary Fig. 1). A similar ambiguity in the binding specificity of group III domains has been reported previously¹². Finally, a minor fraction of the total natural sequences tested (2 out of 28) bound to pSP- and PPLP-

oriented libraries, and were scored as group IV and group II, respectively.

Artificial (CC) WW domains are functionally indistinguishable from natural WW sequences (Fig. 3). Of the ten artificial domains studied, six displayed group I specificity (one is Im) and four showed group III specificity, numbers that approximately reflect the distribution of these sequences in our MSA and in prior samplings of WW specificity^{13,14}. Quantitative binding isotherms measured for all group I artificial sequences show a range of binding affinities typical for natural WW domains (Fig. 3, rightmost column). To probe further the similarity between natural and designed WW sequences, we carried out a saturation mutagenesis study of the group I and group III peptide ligands. In this experiment, every position of the target peptide is replaced individually to each of the 20 amino acids, and the effect on binding is measured by detecting protein hybridization to the array of peptide variants¹⁴. Supplementary Fig. 2 shows these data for two natural WW domains (N39, group I (ref. 10), and N31, group III (refs 5, 14)), and two artificial domains (CC43, group I, and CC20, group III), demonstrating that the designed sequences show a pattern of sensitivity to ligand mutagenesis similar to that for natural sequences. Thus, the amino acid composition at sites plus the information contained in the SCA matrix encodes the sequence rules for quantitatively specifying function in the WW domain.

No artificial sequences with group II or group IV specificity were identified. However, the paucity of group II and group IV sequences

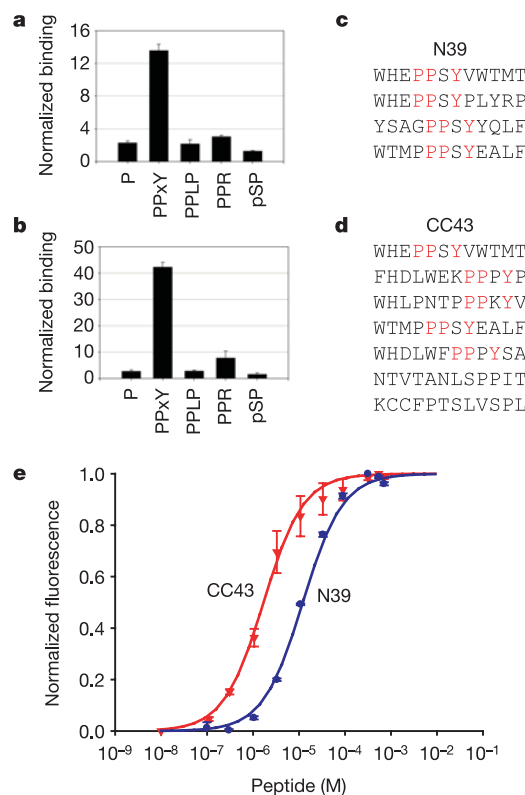


Figure 2 | Assays for binding affinity and specificity in WW domains.

a, b, Oriented peptide library screening (see Methods) of binding specificity for N39, a natural domain with group I (PPxY) specificity (**a**), or CC43, an artificial WW domain showing 37% average and 68% top-hit identity to natural WW domains in the MSA (**b**). Binding is reported as fold above background in the absence of target peptides. **c, d**, Phage display analysis of binding specificity for N39 and CC43 WW domains; both domains select PPxY-containing sequences from a random 12-mer peptide library. **e**, Binding isotherms for N39 (K_d = $11.2 \pm 1.2 \mu\text{M}$) and CC43 (K_d = $1.7 \pm 0.1 \mu\text{M}$), assayed by Trp fluorescence quenching using a model group I peptide (EYPPYPPYPSG). Error bars represent standard deviation from at least four independent assays.

in our alignment of natural WW domains and the relatively small number of artificial domains tested so far probably account for the lack of these specificity classes. These classes may emerge from larger-scale design of artificial WW sequences.

Given these results, can we infer which positions constitute the sequence determinants for WW binding specificity? The eight positions comprising the primary cluster of coevolving residues in the WW family (Fig. 1c) are highlighted in yellow in Fig. 3. Sequences experimentally found to display group I binding specificity, whether natural or designed, strictly conserve a specific sequence motif in the coevolving cluster (³L⁴P⁶G⁸E²¹I/V²²D/N²³H²⁸T). In contrast, the average sequence identity in non-cluster positions is barely different for group I sequences in comparison with all WW sequences: 40.9 ± 2.9% within group I sequences and 35 ± 3.9% overall (mean ± s.d.). These results strongly suggest that this network of mutually evolving residues is the major determinant of group I binding specificity. In accord with this conclusion, marginal group I domains or the weakly specific group III domains display considerable variation within this cluster. A larger study of designed sequences may help to distil the minimal sequence profiles of these and other WW groups.

The spatial organization of the network residues provides an unexpectedly distributed picture of binding specificity in the WW domain. Rather than being restricted to the ligand-binding surface, the eight network residues are organized into a physically contiguous network linking the primary specificity determining pocket (positions 21, 23 and 28) with residues on the opposite side (3 and 4) through a few intervening residues (6, 8 and 22) (Fig. 4a). The

coevolution of these positions predicts that some residues act at long range in mediating peptide binding, and the network amino acids act cooperatively in determining the binding free energy. Previous mutagenesis studies already suggest that network residues at or close to the peptide binding surface (8, 21, 23 and 28) mediate binding specificity^{15–18}, and structural work in the dystrophin WW domain provides a mechanistic understanding for the contribution of these residues in group I domains¹⁹. To test more of the network for contribution to peptide binding, and to test the prediction of cooperative action, we carried out thermodynamic double mutant cycle analysis^{20,21} to measure the energetic coupling between mutations at binding-site position 28 and positions 3, 8 and 23 in the Nedd4.3 WW domain. In the mutant cycle method, the effect of one mutation on the equilibrium dissociation constant for peptide binding is measured in two conditions: (1) the wild-type background ($X1 = K_d^{M1}/K_d^{WT}$), and (2) in the background of a second mutation ($X2 = K_d^{M1,M2}/K_d^{M2}$) (Fig. 4b). The ratio of these effects gives the coupling parameter Ω , a measure of the degree of interaction between the two mutations. If the two mutations are thermodynamically independent, the effect of the first mutation is the same in conditions 1 and 2, and $\Omega = 1$. If $\Omega \neq 1$, then the two mutations act cooperatively.

Consistent with earlier studies^{17,19}, the E8A, H23A and T28A mutations all affected binding of a PPxY-containing peptide (Fig. 4c). However, L3A also had a significant effect (5.15 ± 0.99 fold; mean ± s.d.), although located on the opposite surface from the peptide-binding site. In addition, mutant cycle analyses for the T28A mutation with each of the three other mutations show Ω values

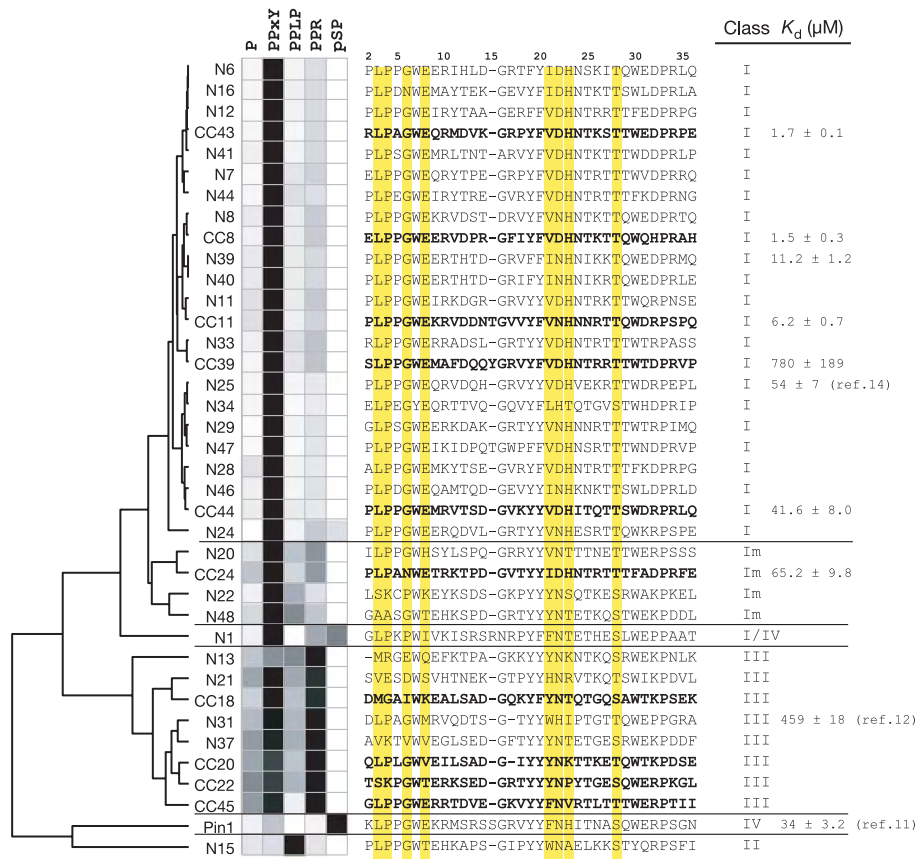


Figure 3 | A summary of functional measurements for all tested natural and artificial WW domains. The matrix shows the results of oriented peptide library assays, clustered to reveal the similarities in binding specificity for different domains. The colour scale ranges from white (minimum signal) to black (maximum signal) for each domain. Class assignments for domains are shown at the right. Dissociation constants for selected group I domains

were either measured by fluorescence binding assay (N39 and all artificial sequences) or were derived from literature references. Sequences are aligned per the MSA, and positions corresponding to the cluster of eight coevolving residues (Fig. 1c) are highlighted in yellow. SCA-based designed sequences are shown in bold. Measurements are mean ± s.d.

that significantly differ from unity (Fig. 4c; see also Supplementary Fig. S3). Specifically, the effects of mutations at 3, 8 and 23 are either diminished (L3A and H23A) or abrogated (E8A) in the background of T28A (Supplementary Fig. 3). Thus T28A is thermodynamically coupled to mutations at 3, 8 and 23. These results support the model that a distributed and cooperative network of residues predicted by the SCA contributes to peptide recognition in the WW domain.

It is perhaps surprising that all folded artificial WW domains could be classified into a known functional group. The SCA emphasizes the deeply conserved couplings between sequence positions in a protein family while down-weighting or even ignoring less conserved couplings. These weak couplings typically arise from small clades of more recently diverged sequences in the MSA, which are expected to contain many positional correlations yet to be relaxed through variation. If these recent branchings of the phylogenetic tree also hold important information about the physical chemistry of specific binding in extant proteins, we might have expected many folded but functionally undifferentiated WW domains in SCA-based design. The data here suggest that at least as defined by *in vitro* assays, information specifying binding specificity is captured in positional interactions that are in the deep evolutionary record. Future studies of functional complementation *in vivo* will help further to address the completeness of the SCA-based sequence information in specifying natural-like proteins.

Compared with the current field of protein design, the SCA-based protein design takes a completely different but complementary approach to understanding the design of natural proteins. Using atomistic energy functions that approximate the physical forces between atoms, several groups have now achieved spectacular successes in the re-design of natural folds^{22–24}, including WW domains²⁵,

and even in the *de novo* construction of artificial folds^{26,27}. These successes demonstrate the accuracy of the scoring functions used, but important unsolved issues remain. The basic design principle in atomistic protein design is optimization of a target potential function that produces sequences with deep thermodynamic minima in the native state and structures with high thermal stabilities^{22,27}. However, natural proteins are thought to have native states with shallow energy minima, resulting in marginally stable but dynamic folds that are often capable of supporting more than one conformation. In a purely statistical and mechanism-free way, SCA-based protein design produces sequences that show the same marginal stability of natural proteins and function like natural proteins. It may be interesting to combine the mechanistic description of energetics from atomistic design with the sparse and distributed architecture of amino acid interactions in SCA to understand better the design of evolved proteins.

METHODS

Statistical coupling analysis. SCA was conducted as previously described^{3,4} on an alignment of 292 WW domain sequences, updated from the original alignment of 120 sequences in the accompanying paper using the March 2004 release of the non-redundant database. SCA results are essentially identical for the two alignments. Sequences were collected using PSI-BLAST (*e*-scores <0.001) and aligned using ClustalW²⁸ followed by manual adjustment. The alignment is available from our laboratory website (<http://www.hhmi.swmed.edu/Labs/rr>), and the code is available upon request.

Oriented peptide library assay. Five biotinylated degenerate peptide libraries were synthesized using N- α -Fmoc protected amino acids and standard BOP/HOBt coupling chemistry. The libraries were constructed to present either a proline-only control (biotin-Z-GMAxxxPxxxAKKK) or the four different characteristic WW domain binding motifs: group-I-oriented (biotin-Z-GMAxxxPPxYxxxAKKK-C), group-II-oriented (biotin-Z-GMAxxxPPLPxxxAKKK), group-III-oriented (biotin-Z-GMAxxxPPRxxxAKKK) and group-IV-oriented (biotin-Z-GMAxxxSPxxxAKKK), where Z is 6-amino-hexanoic acid, pS is phosphoserine, and x denotes all amino acids except cysteine. Peptide libraries were immobilized onto pre-washed streptavidin-coated 96-well plates (10 μ g per well) in phosphate-buffered saline plus 0.5% Tween-20 (PBST) at 4 °C for 1 h. Wells were washed in PBST and incubated with GST-WW domains (0.5 μ g) for 2 h, washed, and detected by ELISA using a horseradish peroxidase-conjugated anti-GST antibody (Amersham) at 1:5,000 dilution for 1 h at 4 °C followed by reaction with 3,3',5,5'-tetramethylbenzidine liquid substrate system (Sigma) for 5 min. Absorbance was monitored at 450 nm.

Phage display. Phage display was carried out using a commercial random 12-mer library (PhD Phage Display kit, NEB) per protocols supplied by the manufacturer. Phage isolates were sequenced after three rounds of amplification and nonspecific elution with glycine-HCl, pH 2.2.

Binding assays. Tryptophan fluorescence-based peptide binding assays were conducted on a PTI spectrofluorimeter, monitoring fluorescence emission at 340 nm (excitation at 295 nm). Group I peptide, EYPPPPPPYPSG (Tufts Protein Chemistry Facility), was purified using reverse-phase high-performance liquid chromatography. Binding assays were conducted at 4 °C in buffer A (100 mM TrisHCl, pH 8.0, 100 mM NaCl) with 1 μ M WW domain. WW domains were cloned into the pHIS8-3 vector, expressed and purified as described². The assay follows the fraction of protein bound to peptide by the normalized fluorescence quenching of the Trp residue in the X-Pro binding pocket. Isothermal titration calorimetry measurements were made using a MicroCal VP-ITC calorimeter in buffer A at 4 °C, starting with 50–200 μ M WW domain in the sample cell and titrating 0.5–3 mM of the group I peptide. Data were fit using MicroCal Origin software provided by the manufacturer, using a one-site binding model. The marginal stability of WW domains required correction of apparent dissociation constants to account for the fraction of folded protein at the assay temperature. Assuming a two-state folding reaction, the correction applied was $K_d = K_{d,app} \left(\frac{K_f}{1+K_f} \right)$, where K_f is the equilibrium constant for the folding reaction. For each protein assayed, K_f was calculated using thermal denaturation studies carried out as in the accompanying paper², and assuming a previously reported value for the change in heat capacity for the unfolding reaction²⁹. K_f values were: wild type, 38.8; L3A, 1.65; E8A, 15.7; H23A, 22.7; T28A, 32.0; L3A/T28A, 1.25; E8A/T28A, 8.54; H23A/T28A, 21.5.

Protein hybridization assay. Protein hybridization assays were performed as described¹⁴. Arrays of all single amino acid substitutions for a group I peptide (GTPPPPYTVG) and a group III peptide (PPGPPPRGPPP) were synthesized on

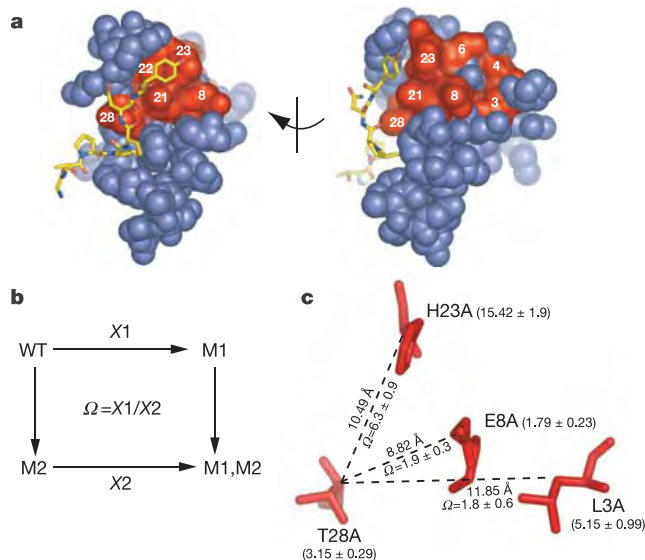


Figure 4 | A spatially distributed network underlying WW function. **a**, The cluster of eight coevolving residues (Fig. 1c) mapped on the Nedd4.3 WW domain structure (in red). The cluster forms a connected network that links binding site residues with the opposite side through a few intervening positions. **b**, The thermodynamic mutant cycle formalism. The fold effect of one mutation (M1) on an equilibrium constant is calculated in the wild-type background (X1) or in the background of the second mutation (X2). The coupling parameter Ω is the ratio of these fold effects (X1/X2); that is, the degree to which the effect of one mutation depends on the second. **c**, Mutant cycle analysis of selected coevolving positions. Residues are shown in the same orientation as in **a** (right), with distances between β -carbons of residues indicated. Single mutations at all sites affect peptide binding (fold effect relative to wild type in parentheses), and mutant cycle analysis demonstrates energetic coupling ($\Omega > 1$) between position 28 and positions 3, 8 and 23. Measurements are mean $\pm$ s.d.

membranes by the M.I.T. Biopolymers Laboratory. Membranes were washed in PBS plus 0.1% Tween-20 (PBST), blocked for 2 h at room temperature in PBST plus 5% non-fat dry milk, washed, and incubated at 4°C overnight with 10–400 µg ml⁻¹ of purified GST–WW domains. Membranes were washed, treated with horseradish peroxidase-conjugated anti-GST antibodies (Sigma) in PBST plus 5% non-fat dry milk at 4°C for 2 h, washed again, and bound WW domains were detected using the ECL kit (Amersham).

Received 25 January; accepted 30 June 2005.

- Voigt, C. A., Kauffman, S. & Wang, Z. G. Rational evolutionary design: the theory of *in vitro* protein evolution. *Adv. Protein Chem.* **55**, 79–160 (2000).
- Socolich, M. *et al.* Evolutionary information for specifying a protein fold. *Nature* doi:10.1038/nature03991 (this issue).
- Lockless, S. W. & Ranganathan, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295–299 (1999).
- Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nature Struct. Biol.* **10**, 59–69 (2003).
- Bedford, M. T., Sarbassova, D., Xu, J., Leder, P. & Yaffe, M. B. A novel pro-Arg motif recognized by WW domains. *J. Biol. Chem.* **275**, 10359–10369 (2000).
- Chen, H. I. & Sudol, M. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc. Natl Acad. Sci. USA* **92**, 7819–7823 (1995).
- Ermekova, K. S. *et al.* The WW domain of neural protein FE65 interacts with proline-rich motifs in Mena, the mammalian homolog of *Drosophila* enabled. *J. Biol. Chem.* **272**, 32869–32877 (1997).
- Lu, P. J., Zhou, X. Z., Shen, M. & Lu, K. P. Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science* **283**, 1325–1328 (1999).
- Zarrinpar, A. & Lim, W. A. Converging on proline: the mechanism of WW domain peptide recognition. *Nature Struct. Biol.* **7**, 611–613 (2000).
- Kanelis, V., Rotin, D. & Forman-Kay, J. D. Solution structure of a Nedd4 WW domain–ENaC peptide complex. *Nature Struct. Biol.* **8**, 407–412 (2001).
- Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T. & Noel, J. P. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nature Struct. Biol.* **7**, 639–643 (2000).
- Kato, Y. *et al.* Common mechanism of ligand recognition by group II/III WW domains: redefining their functional classification. *J. Biol. Chem.* **279**, 31833–31841 (2004).
- Hu, H. *et al.* A map of WW domain family interactions. *Proteomics* **4**, 643–655 (2004).
- Otte, L. *et al.* WW domain sequence activity relationships identified using ligand recognition propensities of 42 WW domains. *Protein Sci.* **12**, 491–500 (2003).
- Chen, H. I. *et al.* Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands. *J. Biol. Chem.* **272**, 17070–17077 (1997).
- Espanel, X. & Sudol, M. A single point mutation in a group I WW domain shifts its specificity to that of group II WW domains. *J. Biol. Chem.* **274**, 17284–17289 (1999).
- Kasanov, J., Pirozzi, G., Uveges, A. J. & Kay, B. K. Characterizing Class I WW domains defines key specificity determinants and generates mutant domains with novel specificities. *Chem. Biol.* **8**, 231–241 (2001).
- Toeptert, F., Pires, J. R., Landgraf, C., Oschkinat, H. & Schneider-Mergener, J. Synthesis of an array comprising 837 variants of the hYAP WW protein domain. *Angew. Chem. Int. Edn Engl.* **40**, 897–900 (2001).
- Huang, X. *et al.* Structure of a WW domain containing fragment of dystrophin in complex with β-dystroglycan. *Nature Struct. Biol.* **7**, 634–638 (2000).
- Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell* **38**, 835–840 (1984).
- Hidalgo, P. & MacKinnon, R. Revealing the architecture of a K⁺ channel pore through mutant cycles with a peptide inhibitor. *Science* **268**, 307–310 (1995).
- Dahiyat, B. I. & Mayo, S. L. *De novo* protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
- Dwyer, M. A., Looger, L. L. & Hellinga, H. W. Computational design of a biologically active enzyme. *Science* **304**, 1967–1971 (2004).
- Kortemme, T. *et al.* Computational redesign of protein-protein interaction specificity. *Nature Struct. Mol. Biol.* **11**, 371–379 (2004).
- Kraemer-Pecore, C. M., Lecomte, J. T. & Desjarlais, J. R. A *de novo* redesign of the WW domain. *Protein Sci.* **12**, 2194–2205 (2003).
- Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T. & Kim, P. S. High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 (1998).
- Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
- Ferguson, N., Johnson, C. M., Macias, M., Oschkinat, H. & Fersht, A. Ultrafast folding of WW domains without structured aromatic clusters in the denatured state. *Proc. Natl Acad. Sci. USA* **98**, 13002–13007 (2001).
- Delano, W. L. The PyMOL Molecular Graphics System (<http://www.pymol.org>) (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Ranganathan laboratory for advice and critical review of the manuscript, J. P. Noel for providing the pHIS8 expression vector, and K. Voegler for contributing to this project. This study was supported by the Robert A. Welch foundation (R.R.), the Mallinckrodt Foundation Scholar Award (R.R.), NIH grants (M.B.Y.), and a Burroughs-Wellcome Career Development Award (M.B.Y.). D.M.L. was supported by a Howard Hughes Medical Institute pre-doctoral award. W.P.R. is an associate and R.R. is an investigator of the Howard Hughes Medical Institute.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.R. (rama.ranganathan@utsouthwestern.edu).

LETTERS

Crystal structure of the RNA component of bacterial ribonuclease P

Alfredo Torres-Larios¹, Kerren K. Swinger¹, Andrey S. Krasilnikov^{1†}, Tao Pan² & Alfonso Mondragón¹

Transfer RNA (tRNA) is produced as a precursor molecule that needs to be processed at its 3' and 5' ends. Ribonuclease P is the sole endonuclease responsible for processing the 5' end of tRNA by cleaving the precursor and leading to tRNA maturation. It was one of the first catalytic RNA molecules identified¹ and consists of a single RNA component in all organisms and only one protein component in bacteria. It is a true multi-turnover ribozyme and one of only two ribozymes (the other being the ribosome) that are conserved in all kingdoms of life. Here we show the crystal structure at 3.85 Å resolution of the RNA component of *Thermotoga maritima* ribonuclease P. The entire RNA catalytic component is revealed, as well as the arrangement of the two structural domains. The structure shows the general architecture of the RNA molecule, the inter- and intra-domain interactions, the location of the universally conserved regions, the regions involved in pre-tRNA recognition and the location of the active site. A model with bound tRNA is in agreement with all existing data and suggests the general basis for RNA–RNA recognition by this ribozyme.

Ribonuclease P (RNase P) processes pre-tRNA, some viral RNAs, 4.5S RNA, tmRNA, C4 antisense RNA from some bacteriophages and a polycistronic pre-messenger RNA². In the case of pre-tRNA maturation, RNase P hydrolyzes the phosphodiester bond immediately 5' of the first nucleotide of mature tRNA, leaving 5' phosphate and 3' hydroxyl termini. Recognition of pre-tRNA by bacterial RNase P involves the acceptor stem, the T ψ C stem³ and the CCA sequence at the 3' end⁴ of the pre-tRNA. The involvement of magnesium in the binding and cleavage of precursor tRNA has been established^{5–8}, and the regions involved in cleavage and recognition have been mapped^{2,9}. Unlike eukaryotic RNase P, in eubacteria and some archaeobacteria the RNA component alone can catalyse pre-tRNA cleavage *in vitro*^{10,11}.

In bacteria, the RNase P RNA component (or P RNA) consists of 300–450 nucleotides. Bacterial P RNAs can be divided into two major types (A or B). Eubacterial P RNAs consist of two independently folded domains: a specificity domain (S-domain), and a catalytic domain (C-domain)¹². The S-domain has been structurally characterized^{13,14}. RNase Ps from all organisms share many sequence similarities in their RNA component, implying that the P RNAs have several universally conserved structural features¹⁵.

The crystal structure of the ancestral, A-type RNA component of *T. maritima* RNase P was solved at a resolution of 3.85 Å. The electron density map allowed tracing of most of the molecule (Fig. 1; see also Methods and Supplementary Information). The molecule is approximately 125 × 90 × 50 Å³ and is made up of two layers, each one-helix thick. Figure 1a depicts a face-on view of the larger of the two layers composed of paired P1–P12 and P15–P17, and including the junctions J5/15, J11/12–J12/11 and loop L15 regions. The second layer

comprises helical stems P13, P14 and P18 (Fig. 1a, lower panel). The two structural domains are visible and interact extensively with each other through P8 and P9 in the S-domain and P1, P4 and P18 in the C-domain. The larger layer of the structure contains most of the universally conserved regions, the areas expected to contact the pre-tRNA substrate and the putative active site (see below). The S- and C-domains represent independent structural units that form extensive inter-domain interactions.

The S-domain structure is largely identical to the one of *Thermus thermophilus* RNase P (ref. 14), another A-type molecule, and underscores the structural stability of the domain. The slightly concave C-domain has several distinctive features: (1) a long stem formed by the co-axially stacked P1, P4 and P5 helices; (2) a second long stem formed by the co-axially arranged P2 and P3 helices; (3) a third stem, P18, which runs almost perpendicular to the other two helices and links P2/P3 to the P15/P17 stem, and is the only large structural element not in the same layer; (4) a stem that folds on itself (P15/P17) and whose end forms a pseudoknot with P6; and (5) a non-canonical region that links these structural elements and is located between P2/P3 and P4.

The P1/P4/P5 and P2/P3 stems are at a ~15° angle from each other, closer to the parallel orientation¹⁶ than the perpendicular orientation¹⁷ which was predicted in earlier models. P18 is linked to P15 and P2 and runs almost perpendicular to the other helices of the C-domain. No canonical tertiary interactions are seen between the helices forming the C-domain, but the four major structural elements (P1/P4/P5, P2/P3, P15/P16 and P18) are held together by linkers that include several highly conserved nucleotides that form complex interactions.

The molecule crystallized as a dimer via formation of an intermolecular version of the pseudoknot P6 (see Supplementary Information). From the structure, the functional significance of this dimer cannot be assessed, but an intramolecular P6 pseudoknot (Fig. 1) can be considered by including coordinates for P16/P17/P6 from a symmetry-related molecule; although this places P15 and P16 too far apart to interact through the short L15 internal loop. For intramolecular pseudoknot formation, there has to be a dramatic kink in the P15/P16/P17 helices around the L15 region, which has previously been suggested¹⁶. The intermolecular pseudoknot also indicates that the P15/L15/P16 region is dynamic and that the orientation of P15 and P16/P17/P6 could easily change, bringing L15 closer to the main body of the molecule. In subsequent discussions, the intramolecular pseudoknot is assumed.

The structure of the C-domain is formed mostly by helices, with several non-helical regions joining them. At the centre of the P1/P4/P5 helix there is a region where several of these non-helical regions converge. The nucleotides joining P2/P3 to P4 at the P1/P4 junction and P2 to P18 at the P4/P5 junction come close in space and

¹Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208, USA. ²Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA. [†]Present address: Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

face the outside of the molecule. These nucleotides are all in the same general region, filling the space between the P1/P4/P5 and P2/P3 helices. The nucleotides linking P15 to P18 and P15 to P5 are also in the vicinity. At the resolution of the structure it is not possible to assign a position for these bases, but the phosphate backbone could be traced with confidence.

P3 is the only element in the C-domain that is far from the core of the molecule. Its role could be to link the universally conserved regions (Fig. 2). P3 can be replaced and shortened but not deleted¹⁸. P1 is essential for the proper folding of the P RNA due to its interaction with P9¹⁹. Both P4 and P18 are essential, with P18 exhibiting covariation with P8 and P14²⁰. The role of P4 has been

extensively analysed due to its sequence conservation, proximity to the catalytic site and its potential to bind the metal ions involved in catalysis. Nevertheless, the structure suggests that its role could be structural and its sequence conservation could be due to constraints imposed by its proximity and interaction with the connectors in the C-domain.

The molecule forms a concave structure with the S- and C-domains intimately interacting with each other. The central organizer of the domains is the P8/P9 stack. Tetraloop–tetraloop receptor interactions occur between P8 and L14, P8 and L18 and L9 and P1 (ref. 21). An A-minor interaction is found between L8 and P4 (see Supplementary Information). All these interactions ensure that the

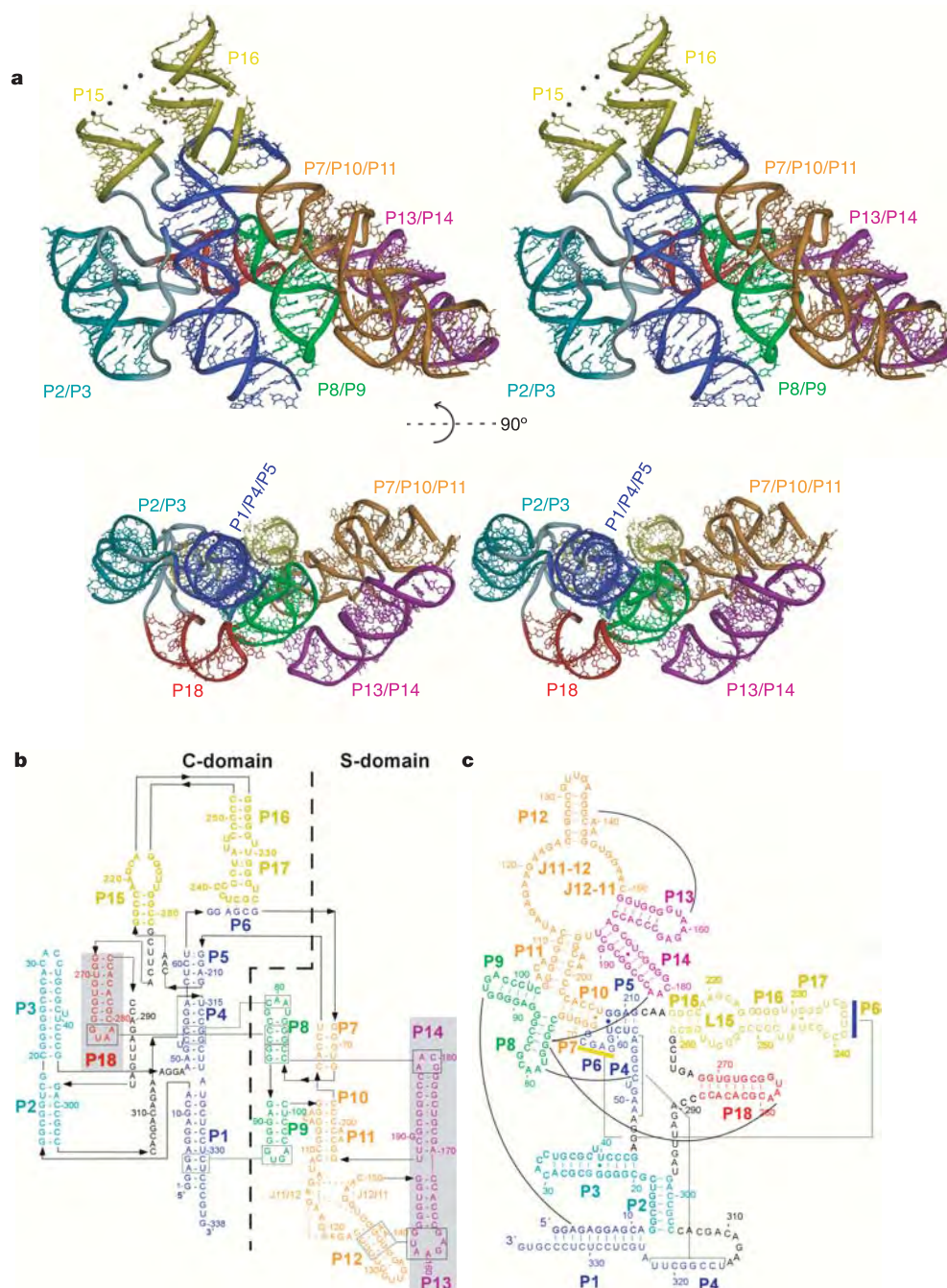


Figure 1 | Schematic diagram of *T. maritima* RNase P RNA. **a**, Stereo ribbon diagram of two orthogonal views of the structure. **b**, **c**, Sequence and secondary structure diagrams of the P RNA showing major interactions observed in the crystal structure. Filled circles represent non-canonical base

pairing; dashed lines correspond to other types of interactions; arrows indicate 5' to 3' direction; and lines linking boxed nucleotides represent tertiary interactions. Shaded boxes correspond to regions in the smaller of the two layers.

two domains are positioned correctly and oriented in the manner appropriate for tRNA binding. On the tertiary structure level, interactions between hairpin loops and helices dominate. Overall, all the interactions between domains predicted by a variety of approaches are observed, and the structure agrees well with earlier models and with secondary structure predictions^{16,17}.

Sequence analysis suggests that all P RNAs have a common core that includes five conserved regions (CR I–V; refs 15, 22) (Fig. 2a; see also Supplementary Information). CR II and III are located in the J11/12–J12/11 module of the S-domain. This region has an unusual fold and contains the highly conserved G147 involved in tRNA recognition^{23,24}. CR IV is located in the linker joining stems P2 to P18 (Fig. 2), in between P2/P3 and P1/P4/P5. The CR IV region neighbours other highly conserved nucleotides and it may have a structural role by helping to confer a specific organization to the active site. The two other conserved regions, CR I and CR V, are also non-helical and are clustered together and located adjacent to the P4 stem. The presence of several conserved regions of nucleotides lacking an easily identifiable secondary or tertiary structure suggests that sequence conservation may reflect structural constraints.

Substrate recognition involves contacts with nucleotides in the S-domain (in P4 and around P5), through base pairing of the CCA-3' end with nucleotides in the L15 loop, and with the protein com-

ponent. The interacting nucleotides in the S-domain have been mapped^{23,25}, and two correspond to G147 and A198 in *T. maritima*. In the case of *Bacillus subtilis* RNase P (refs 24, 25), these nucleotides interact with nucleotides 54, 56, 61 and 62 in the T Ψ C stem of tRNA, and it is likely that the interactions are conserved in all P RNAs. The CCA at the 3' end of the pre-tRNA forms base pairs with nucleotides around the L15 internal loop⁴. Additionally, interactions have been mapped to the conserved regions in P4, the J5/15 linker and the J2/4 connector (summarized in refs 2, 9) (Fig. 2b; see also Supplementary Information).

To gain a better understanding of pre-tRNA recognition, tRNA was modelled into the structure (Fig. 3). In this model, the T Ψ C stem loop lies in the S-domain opening and the 5' end of the tRNA lies near the P4/P5 region. The acceptor stem of the tRNA sits on the shallow groove created by the concave C-domain. The tRNA stem runs almost parallel to P1/P4/P5 and passes near the conserved nucleotides in this region. The cleavage site sits near the universally conserved residues A49 in P4 and A314 in J2/4. The cleavage site is also near J5/15, although some changes would be required to form interactions between this region and the tRNA. J5/15 contains A213, which is highly conserved in all bacteria and implicated in specific interactions with tRNA²⁶. A213 is found between the region involved in base pairing with the CCA at the 3' end⁴ and the conserved regions around P4. Conserved U52 in P4 forms an unstacked base that cannot be seen in the map. In our model the phosphate of U52 is found near the tRNA, but on the other side of the stem from where the conserved nucleotides are located. Thus it seems that A49, A213 and A314 mark the general location of the active site. Finally, in the tRNA/RNase P model, the unpaired CCA at the 3' end of tRNA is close to the L15 region. The L15 loop would have to move in order to interact with the pre-tRNA substrate. The fact that the P15/L15/P16

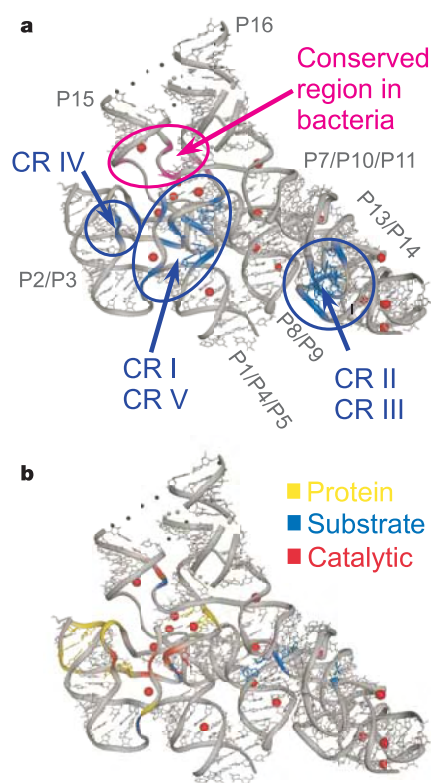


Figure 2 | Conserved and interacting regions in RNase P. **a**, Universally conserved regions¹⁵ are shown in blue and a highly conserved region in bacterial P RNAs is shown in magenta. CR II and III map to the S-domain. CR IV maps to the P2 to P18 connector. The other regions are located in nucleotides near P4. The role of these conserved regions appears to be structural, as they are found in areas that have a well-defined conformation. A highly conserved region in bacteria near A312 is located at the J5/15 junction and has been implicated in direct contacts with the substrate²⁶. This region, together with the vicinity of the universally conserved nucleotides A49 and A314, may correspond to the general location of the active site. **b**, The nucleotides that are involved in interactions with the substrate (blue) or the protein (yellow) are shown and those that are involved in catalysis are shown in red (see Supplementary Information; and refs 2, 9, 27). Red spheres are osmium atoms.

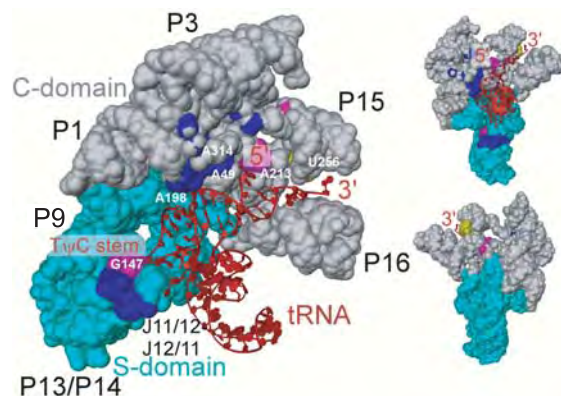


Figure 3 | Model of RNase P-tRNA interactions. The diagram shows the accessible surface area of P RNA and a diagram of tRNA^{Phe}. The model was built using the higher resolution structure of *T. thermophilus* S-domain as a guide, as the interactions between the S-domain nucleotides (G147 and A198) and the tRNA T Ψ C stem are well characterized^{23,25}. In this model, the T Ψ C stem loop lies in the S-domain opening and the 5' end of the tRNA lies near the P4/P5 region and putative active site. The distance between the more distal points in these two regions is around 46 Å, whereas the distance between the T Ψ C stem loop and the cleavage site in mature tRNA^{Phe} is around 41 Å. The 3' CCA sequence is close to the L15 loop, where base pairing interactions occur. Colour coding is as follows: dark blue, nucleotides in the universally conserved regions; pink, highly conserved nucleotides in bacteria; and yellow, the L15 region that interacts with the pre-tRNA CCA 3' end. The views are as follows: the concave side of RNase P (left); down the anti-codon arm of tRNA (top right); and backside of the molecule (bottom right). The general active site location is marked by the 5' end of the tRNA molecule, in the vicinity of universally conserved nucleotides A49, A314 and the highly conserved bacterial nucleotide A213. The model conveys the general arrangement of the two molecules and the excellent fit between the tRNA and the two major regions of interaction with P RNA.

region seems to be quite dynamic indicates that this may be a region where conformational changes occur to accommodate the substrate. The model, although built independently, has an overall similarity to an RNA/holoenzyme model²⁷, suggesting that the holoenzyme model captures the general location of the protein correctly.

RNA–RNA recognition involves the presence of a structurally well defined and conserved specificity site that is responsible for selecting tRNA molecules through interactions with the T ψ C stem loop. The catalytic active site appears to be fully formed even in the absence of the pre-tRNA substrate. Other RNA–RNA interactions appear to be non-specific and may involve metal ions. They concentrate along the acceptor stem of tRNA and the P4 stem of RNase P. Thus, RNA–RNA recognition involves regions that make specific contacts with the substrate, a wide range of non-specific interactions between the two molecules, and a fully or almost fully assembled active site. A fuller understanding of tRNA recognition and RNase P catalysis awaits further structural work at atomic resolution.

METHODS

The crystal structure of the 338-nucleotide-long *T. maritima* RNase P molecule was solved by SIRAS using osmium hexamine and cobalt hexamine soaked crystals. The structure was phased with SHARP²⁸ and refined with Refmac5 (ref. 29) and CNS³⁰. Two temperature factors were assigned per nucleotide. Due to the resolution of the data, no attempts were made to place nucleotides not belonging to helical regions, except in the S-domain, where the structure was built using the *T. thermophilus* structure as a guide. Of the 338 nucleotides in the molecule, 309 were built. For nucleotides 39, 45–48, 52, 62, 68, 207, 212–214, 262–267, 288–297 and 305–314 only the phosphate backbone was built. Density was not found for nucleotides 30–33, 120–122, 131–135, 221–223, 229, 234, 239–241, 247–248, 253–255 and 335–338. The final R_{factor} and R_{free} for the structure were 35% and 36.6% respectively. Figures were prepared with Dino (<http://www.dino3d.org>). For more complete details see the Supplementary Information.

Received 15 June; accepted 28 July 2005.

Published online 21 August 2005.

- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
- Christian, E. L., Zahler, N. H., Kaye, N. M. & Harris, M. E. Analysis of substrate recognition by the ribonucleoprotein endonuclease RNase P. *Methods* **28**, 307–322 (2002).
- McClain, W. H., Guerrier-Takada, C. & Altman, S. Model substrates for an RNA enzyme. *Science* **238**, 527–530 (1987).
- Kirsebom, L. A. & Svard, S. G. Base pairing between *Escherichia coli* RNase P RNA and its substrate. *EMBO J.* **13**, 4870–4876 (1994).
- Christian, E. L., Kaye, N. M. & Harris, M. E. Evidence for a polynuclear metal ion binding site in the catalytic domain of ribonuclease P RNA. *EMBO J.* **21**, 2253–2262 (2002).
- Beebe, J. A., Kurz, J. C. & Fierke, C. A. Magnesium ions are required by *Bacillus subtilis* ribonuclease P RNA for both binding and cleaving precursor tRNA^{Asp}. *Biochemistry* **35**, 10493–10505 (1996).
- Perreault, J. P. & Altman, S. Pathway of activation by magnesium ions of substrates for the catalytic subunit of RNase P from *Escherichia coli*. *J. Mol. Biol.* **230**, 750–756 (1993).
- Smith, D. & Pace, N. R. Multiple magnesium ions in the ribonuclease P reaction mechanism. *Biochemistry* **32**, 5273–5281 (1993).
- Kurz, J. C. & Fierke, C. A. Ribonuclease P: a ribonucleoprotein enzyme. *Curr. Opin. Chem. Biol.* **4**, 553–558 (2000).
- Pannucci, J. A., Haas, E. S., Hall, T. A., Harris, J. K. & Brown, J. W. RNase P RNAs from some Archaea are catalytically active. *Proc. Natl Acad. Sci. USA* **96**, 7803–7808 (1999).
- Reich, C., Olsen, G. J., Pace, B. & Pace, N. R. Role of the protein moiety of ribonuclease P, a ribonucleoprotein enzyme. *Science* **239**, 178–181 (1988).
- Loria, A. & Pan, T. Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA* **2**, 551–563 (1996).
- Krasilnikov, A. S., Yang, X., Pan, T. & Mondragón, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**, 760–764 (2003).
- Krasilnikov, A. S., Xiao, Y., Pan, T. & Mondragón, A. Basis for structural diversity in homologous RNAs. *Science* **306**, 104–107 (2004).
- Chen, J.-L. & Pace, N. R. Identification of the universally conserved core of ribonuclease P RNA. *RNA* **3**, 557–560 (1997).
- Massire, C., Jaeger, L. & Westhof, E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* **279**, 773–793 (1998).
- Harris, M. E. *et al.* Use of photoaffinity crosslinking and molecular modeling to analyse the global architecture of ribonuclease P RNA. *EMBO J.* **13**, 3953–3963 (1994).
- Guerrier-Takada, C. & Altman, S. Reconstitution of enzymatic activity from fragments of M1 RNA. *Proc. Natl Acad. Sci. USA* **89**, 1266–1270 (1992).
- Harris, M. E., Kazantsev, A. V., Chen, J. L. & Pace, N. R. Analysis of the tertiary structure of the ribonuclease P ribozyme-substrate complex by site-specific photoaffinity crosslinking. *RNA* **3**, 561–576 (1997).
- Brown, J. W. *et al.* Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements. *Proc. Natl Acad. Sci. USA* **93**, 3001–3006 (1996).
- Massire, C., Jaeger, L. & Westhof, E. Phylogenetic evidence for a new tertiary interaction in bacterial RNase P RNAs. *RNA* **3**, 553–556 (1997).
- Siegel, R. W., Banta, A. B., Haas, E. S., Brown, J. W. & Pace, N. R. Mycoplasma fermentans simplifies our view of the catalytic core of ribonuclease P RNA. *RNA* **2**, 452–462 (1996).
- LaGrandeur, T. E., Huttenhofer, A., Noller, H. F. & Pace, N. R. Phylogenetic comparative chemical footprint analysis of the interaction between ribonuclease P RNA and tRNA. *EMBO J.* **13**, 3945–3952 (1994).
- Odell, L., Huang, V., Jakacka, M. & Pan, T. Interaction of structural modules in substrate binding by the ribozyme from *Bacillus subtilis* RNase P. *Nucleic Acids Res.* **26**, 3717–3723 (1998).
- Loria, A. & Pan, T. Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry* **36**, 6317–6325 (1997).
- Zahler, N. H., Christian, E. L. & Harris, M. E. Recognition of the 5' leader of pre-tRNA substrates by the active site of ribonuclease P. *RNA* **9**, 734–745 (2003).
- Tsai, H. Y., Masquida, B., Biswas, R., Westhof, E. & Gopalan, V. Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme. *J. Mol. Biol.* **325**, 661–675 (2003).
- de la Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Brunker, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Yusupov, B. Golden and J. Cate for their gift of osmium and iridium hexamine, and Y. Zhang for mass spectrometry. We thank A.-C. Dock-Bregeon, E. Sontheimer and O. Uhlenbeck for comments and suggestions, P. Nissen for advice and A. Vega-Miranda for help with the figures. Research was supported by an NIH grant to A.M. Support from the R.H. Lurie Cancer Center of Northwestern University to the Structural Biology Center is acknowledged. Portions of this work were performed at the DND, LS, IMCA and SER Collaborative Access Teams at the Advanced Photon Source. We thank members of these teams for their help and support. DND-CAT is supported by DuPont, Dow, and the NSF, and use of the Advanced Photon Source is supported by the DOE.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 2A2E. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M. (a-mondragon@northwestern.edu).

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

An infinite learning curve

Writer David Foster Wallace's most famous works, *Everything and More* and *Infinite Jest*, are concerned literally and metaphorically with infinity. Wallace sometimes describes himself as "ridiculously overeducated" — a description that many scientists, with years of postgraduate education and postdoctoral training under their belts, will recognize. Yet, like Wallace, these scientists still feel the need to learn more.

Of course, this desire can be sated by simply picking up a book to explore a different discipline. But in recent years more formal mechanisms for expanding education have begun emerging. These are helping scientists learn how to work across disciplines and often provide career workshops to give them off-the-bench skills.

One programme, run by the US National Institutes of Health, trains chemists, engineers, physicists and mathematicians in biological research. Since it launched in 2000, more than 120 mid-career scientists have gone back to the classroom to pursue years of additional coursework. Members of its first class are now ready to combine their physical background with a newfound

biological education, which should make them more effective interdisciplinary researchers (see page 470).

Another programme, run by the US-based Committee on the Advancement of Women Chemists (see page 592), offers workshops to help female chemists and engineers leap gender hurdles. These courses emphasize communication, leadership and negotiation skills. So far the workshops have proved to be a big success with participants and, as a result, the organization is considering expanding its reach to Europe.

The interest in both of these programmes shows that there is nothing ridiculous about education beyond the PhD and postdoc. Whether skills are scientific or off-the-bench, in today's competitive environment, the need to learn more seems to go on and on, ad infinitum.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York
 345 Park Avenue South, 10th Floor,
 New York, NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
 Peter Bless
 Classified Sales Representatives
 Tel: +1 800 989 7718

**New York/Pennsylvania/
 Latin America:** Kelly Roman
**Midwest USA/Maryland/
 NIH:** Wade Tucker
East USA/Canada:
 Janine Taormina

San Francisco Office
Classified Sales Representative:
 Michaela Bjorkman
 225 Bush Street, Suite 1453
 San Francisco, CA 94104
 Tel: +1 415 781 3803
 Fax: +1 415 781 3805
 e-mail: m.bjorkman@naturesf.com

European Head Office, London
 The Macmillan Building,
 4 Crinan Street,
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director: Nevin Bayoumi (4978)
European Sales Manager: Andy Douglas (4975)

Advertising Production Manager: Billie Franklin
 To send materials use London address above.
 Tel: +44 (0) 20 7843 4814
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
 Patrick Phelan
 e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

TAG TEAMS

A collaboration can produce powerful results when everyone pulls together, but if you go about it the wrong way, or with the wrong people, it may all fall down around you.

Kendall Powell finds out how to choose the right partners.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Daniel Rizzuto and his postdoctoral adviser study motor control in patients who have electrodes implanted in their brains. As neither is medically qualified, they have to collaborate with neurosurgeons and neurologists.

Crossing the academic-clinical divide, the project hit a major roadblock when ethical considerations ruled out one type of surgery as too risky. Luckily, Rizzuto's partnership at the California Institute of Technology in Pasadena, held two key ingredients to overcome obstacles — good chemistry between the collaborators and open lines of communication. The team eventually settled on a project that met patient guidelines and research goals. "It took a different turn from what we had anticipated, but it's been very fruitful," says Rizzuto.

An unexpected twist is just one of the advantages of collaborating. It can also open up funding opportunities, raise a researcher's visibility and reveal fresh ways to tackle problems. Collaborations come in all varieties — small or multi-group, on campus or between campuses, with big names or with junior peers. New investigators must weigh the benefits of collaboration against their institution's expectations for tenure and the time and resources it may take from core independent research. If lab heads decide to collaborate, they should learn from past efforts the keys to smooth operations — and how to bail out gracefully when collaborations turn sour.

A helping hand

For pre-tenure faculty members thinking of teaming up, the most important consideration is how their department will view collaborative research. Obviously, rookie principal investigators must first establish their own core research programme before branching out. But opinions vary as to whether group work adds to or distracts from that core research.

"There is a perception in some departments that collaboration is a sign of weakness," says Federico Rosei, a nanomaterials physicist at the University of Quebec's National Institute for Scientific Research near Montreal, Canada. But at the same time, he says, collaborations have become increasingly important as scientists now tend to over-specialize in narrow fields.

With expertise in single-molecule measurements, Tom Perkins, an assistant professor at the University of Colorado's physical-sciences institute JILA in Boulder, has several ongoing collaborative projects. Because the work spans diverse disciplines, he says, the individual contributions are clear and no one rides on anyone else's coat-tails. But if there is any doubt, he says, "you want to talk to your chair and understand how your university is going to see this collaboration".

Perkins also notes that team efforts, especially interdisciplinary ones, can provide alternative sources of funding for a beginner. For example, the Human Frontier Science Program (for applicants from much of Europe, Japan, South Korea, Australia and the United States) and the US National Academies' Keck Futures Initiative are "giving out rather large hunks of money to bring people with expertise together", Perkins says.

Small seed grants for collaborations can generate preliminary data for later individual grant applications. But again, a young investigator should know how collaborative grants count in tenure calculations, compared with grants made to a single investigator.

Special cases can also make or break a young collaborator's decision. Steer clear of projects that require international fieldwork or other situations with too many uncontrolled variables. A unique technology can be a boon or a bane — collaborations may be necessary to win institutional support for high-tech, high-cost equipment, but could also lead to getting bogged down with requests.

The right people

There are two main options when it comes to selecting a collaborator: stick to colleagues on your level or shoot for the stars in a field. Rosei warns that collaborating with big shots is a mistake.

"They are busy and will rarely be as committed as you are," he says. Plus, they may get most of the credit even if you do the lion's share of work. Priority and responsibility "will never be equal," he notes.

Tasha Belfiore, a postdoc at the University of California, Berkeley, who has worked on group projects, is especially sensitive to unequal partnerships. "At an early stage, you have a lot more at stake if someone sits on a manuscript or is acting like a jerk," she says. She had better experiences with colleagues "at my own level, where I have more leverage".

But Taekjip Ha, a biophysicist at the University of Illinois at Urbana-Champaign, takes the opposite view. "If you want to work on something that you don't have expertise in, then look for the best person in the world," he says. Having the world's expert in your camp will give the work credibility and ease publication, says Ha. "That's always preferable to working on your own."

Also, Ha adds, well-funded senior collaborators can easily divert resources or postdocs to a project.

If you're a young researcher who joins a group with established, senior investigators, you can jump-start your career, says Simon Watkins, director of the Center for Biologic Imaging at the University of Pittsburgh. "The others will talk about you and raise your visibility — that's important in the hunt for tenure."

Regardless of seniority and reputation, look for personality traits in collaborators such as reliability, honesty and openness. Search out researchers who have the same level of interest, enthusiasm and commitment to pursuing the question at hand. Perkins always begins collaborations with a personal contact by phone or at a meeting, then discusses the idea over several weeks to gauge the other's interest.

Meeting of minds

Most importantly, find someone you 'click' with, says Rizzuto: "If you don't enjoy working with them, how can anything good come of it?" Experienced collaborators say that members of a partnership with complementary knowledge and skills sets will avoid stepping on each other's toes. And those interactions tend to be the most productive.

Catherine Baty, a biological imager at Pittsburgh, says that having collaborators who are businesslike and well organized makes her job easier, but cautions that 'easier' does not always mean 'better'. "Some of my collaborators make me want to take Valium, and yet they are really productive," she notes. "A person who is creative in a different way can make for a potentially powerful collaboration."

The basic recipe for glitch-free collaboration boils down to straightforward communication. The first



S. MYOONG



Tips from team players, from top: Taekjip Ha always speaks to collaborators rather than e-mailing; Simon Watkins advises joining established researchers; Daniel Rizzuto says you have to find people you 'click' with; Tom Perkins says check how your university sees your project first.

meeting — ideally conducted in person — should spell out expectations, rules and responsibilities of the collective effort as well as how publication authorship and intellectual property will be divided. Keep all shared data private until the agreed time for disclosure. To ensure no one is left out of the loop, conference calls and e-mails should include everyone involved.

Ha phones colleagues once a month or quarter to be sure projects stay on track and queries get answered. When possible, he invites off-campus collaborators to visit and "tries to feed them well". He also encourages student and postdoc exchanges to learn or at least appreciate new techniques and their limitations.

The written communication of proposals or end results is equally important. Perkins uses online meeting software with his co-authors to edit manuscripts in real-time on the Internet. And the physics-oriented Ha relies heavily on his biochemist partners to keep him from "writing things that are plain stupid, and that would offend reviewers", he says.

When a problem occurs, use the phone and not e-mail, warns Ha, to avoid misunderstandings or delays. If you are at fault, apologize profusely, remedy the situation quickly and move on. Perkins notes that having separate accounts for shared money is a simple way to avoid budget conflicts. And, Watkins points out, on-campus collaborators can always be hunted down on foot for accountability if need be.

Team spirit

If frustrations arise in a large group collaboration, a young investigator should have an ally in the lead investigator, whose role includes that of enforcement officer when someone neglects their duties.

But even with the best intentions, not every collaboration will be a perfect match. "Identify a bad collaboration and weed it out as soon as possible," urges Rosei. Otherwise, it will drag down your other research, siphoning off too much time and energy.

Science becomes more collaborative every day, so young scientists should remember that playing well with others will never go out of style. Team science has its rewards, too. Perkins enjoys learning to speak the disciplinary language of his collaborators and broadening his scientific approach.

He is not stressed about how his collaborations will reflect in his tenure decision. "Fun, interesting science in my field is usually very collaborative," he says. "I'm much more worried about the quality of science I get done, rather than who's on the author list. And then, you just hope your colleagues appreciate getting quality science done."

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

WEB LINKS

HHMI/BWF course, Chapter 12 on collaborating
 • www.hhmi.org/grants/pdf/labmgmt/ch12.pdf
 Human Frontier Science Program
 • www.hfsp.org
 Keck Future Initiatives
 • www7.nationalacademies.org/keck

MOVERS

Stanley Plotkin, board of directors, Dynavax, Berkeley, California



1999-present: Executive adviser, Sanofi Pasteur, Lyons, France

1991-98: Medical and scientific director, Pasteur Mérieux Connaught, France

1991-present: Professor emeritus of paediatrics, University of Pennsylvania

1991-present: Professor emeritus of virology, Wistar Institute, Philadelphia, Pennsylvania

Stanley Plotkin is widely recognized as a leading authority in vaccinology, and is adviser to the chief executive of vaccine manufacturer Sanofi Pasteur in Lyons, France. Last month, he also took on his very first position on the board of directors for a company — at Dynavax Technologies in Berkeley, California.

The path to his success began when he joined the Epidemic Intelligence Service at the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia. After his training, he angled for an assignment away from epidemiology in the anthrax laboratory at the Wistar Institute in Philadelphia, Pennsylvania. Although he was not particularly interested in anthrax, Plotkin says he took “a gamble” so that he could work with Hilary Koprowski, who created the first live polio vaccine.

It was an outbreak of inhalation anthrax in New Haven, Connecticut, that changed Plotkin's career. With his co-workers on holiday, Plotkin not only had to detail the clinical course and spread of the infection, but also had to develop a preventative vaccine. “By choosing something no one else wanted, anthrax, I was successful in doing what I hoped to do,” he says.

The success of working on a vaccine for a disease that others considered untouchable led Plotkin further away from epidemiology. He subsequently worked on vaccines against polio, varicella, rabies, rotavirus and rubella while at the Wistar Institute and the Children's Hospital of Philadelphia. But one vaccine has eluded him for 25 years: that against cytomegalovirus. The development of a vaccine against a chronic infection such as this presents difficulties not faced with other diseases, he notes. He calls the quest his hobby, but remains optimistic that he will see a licensed vaccine in his lifetime.

Working in both academia and industry has helped Plotkin to guide vaccines from basic research to the marketplace. Having worked as medical and scientific director for Pasteur Mérieux Connaught, Plotkin understands how vaccines are manufactured — often a perspective foreign to academics.

Plotkin advises young scientists eager to develop vaccines to consider working in industry. Despite the ongoing need for basic discoveries, he says that tighter safety and regulatory standards make it easier to develop vaccines in industry than in universities. But the most important thing students can do, he says, is seek the counsel of an adviser who will grant them all the responsibility they can handle. ■

Virginia Gewin

SCIENTISTS & SOCIETIES

Redressing the balance

“I am the only female in my chemistry department and I have often wondered why I have such a hard time being heard in faculty meetings. Sometimes it seems as if I'm being ignored, but then a few minutes later, someone else will say almost my exact words and his comment gets agreement and acceptance.”

This is the voice of one of hundreds of women science and engineering faculty members who have come to workshops in the United States run by the Committee on the Advancement of Women Chemists (COACH).

The COACH workshops aim to eliminate gender-bias barriers that hamper the progress of women in science and, in the meantime, to help women faculty members to move up the career ladder and become leaders in their research, teaching or administrative roles. These successes then create role models and mentors for younger women scientists.

Over the past four years, COACH has run workshops to help women enhance their communication and negotiation skills and gain effective leadership techniques. They have also provided information on creating strategies to make institutional and departmental changes that can improve the climate, recruiting and retention of underrepresented groups.

Case studies, theatre, role-playing

and lively debate contribute to the learning experience for the 15–20 participants in each session.

More than 300 women have attended the workshops in the past four years. As word has spread through the science community, women scientists in other disciplines have hired COACH lecturers to hold workshops at their home institutions and professional meetings, adding another 700 female academic scientists to the tally.

In October, COACH will hold its first workshop specifically geared towards ethnic minority women who are science faculty members and postdoctoral associates. And it is now collaborating with women scientists in Germany, Britain and South America to bring workshops to these countries. The organization is also designing a series of forums for men and women faculty members and administrators, which will be aimed at improving inclusivity, thereby enhancing faculty and department productivity. The popularity of its workshops in the United States and the interest in them from abroad shows that COACH is fulfilling a previously unmet need. ■

Geraldine Richmond, Richard M. and Patricia H. Noyes Professor of Chemistry, University of Oregon, co-founder and chair of COACH.
♦ <http://coach.uoregon.edu>

GRADUATE JOURNAL

Back-up plans

My trip to Alaska this summer provided some perspective — both welcome and unwelcome. After spending my summer on the road, I finally laid eyes on the lights of Los Angeles earlier this month. It was late at night and I was too tired to have any profound feelings. A pillow and a real bed were all I desired, so I parked my car and reached for the only possession in it that just cannot be replaced — my laptop. And it was gone. I lost my mind as I made a mental inventory of the files I had created since my last back-up. Days of writing, all of my trip pictures and unpublished lab work were among the first items to come to mind.

I became overwhelmed with worries about my professional future. I had no computer, no job and dwindling funds. But after the initial panic, I quickly began to regroup. Fortunately, a gas station where I had stopped earlier found my bag and is sending it to me. And I've got a lead on an adjunct teaching position for next semester. I still haven't found a job for the short term, but I am not so worried about that — the trip reminded me that there's more to life than science and a career. Now that I've secured my laptop and have some job leads, my future looks brighter than it did when I left for Alaska. And the trip definitely taught me the importance of back-ups — both in terms of computer data and career plans. ■

Jason Underwood completed his PhD in molecular biology at the University of California, Los Angeles, in June.

Falling

The view from here.

Benjamin Rosenbaum

You're on the 236th-level Kaiserstrasse moving sidewalk when you see her.

You're leaning on the railing, waiting to ask Derya about a job, watching the glittering stream of mites that arc over half the sky — flying up to rewind their nanosprings in the stratospheric sunlight, flying down to make Frankfurt run. You never get tired of watching them.

She's on the Holbeinsteg bridge. Someone's hung it up here — 100 metres of clean grey and green twentieth-century modernism, plucked up from the River Main and suspended in the chilly air 2,360 metres up, between a lump of wooded parkland and a cluster of antique subway cars. She's wearing a 1950s sundress and a broad-brimmed hat, and it's like an essay on the last century — the austere steel bridge, the bright blobs of subway graffiti, and her yellow dress, flapping against her legs as she climbs over the bridge's rail. A picture of elegance and style from the age of money, violence and simplicity.

She's a strawberry blonde, slim, her skin blank and virginal as new butter. She's beyond the rail now, hanging out over the mountain-high drop. Thin translucent shadows move across her, the shadows of the neosilk-and-nanotube filaments that hang the city from the hundreds of 5-kilometre-high towers that encircle it. (A civic agent notices you noticing, and attaches itself to your infospace, whispering statistics — *each object's suspension must weather a class-5 hurricane and the destruction of 80% of the towers, and Frankfurt's current population is stable at 53 million and average age 62, birthrate 0.22, net immigration of half a million a year and current personal squat-right — 311 cubic metres per resident* — until you brush it away.)

You're watching her lean out. The wind whips her hair, ruffles the skirt around her knees. She must be a tourist. You remember your first trip to the upper levels: leaning over the edge into the angry swarm of mites, whirring and buzzing warnings and shoving you back like a million mosquito chaperones. Everyone tries it once...



Except that there aren't any mites around her.

You clutch the railing. Hot, animal fear surges in your chest.

She looks up at you and, across the gap of 40 metres, smiles a brilliant, heart-breaking smile.

Then she lets go and falls.

You scream.

"Bloody airmen," says Derya. He steps off the moving sidewalk near you. Tall, hook-nosed, the fashionable whorls of pox and acne making constellations of his cheeks and chest, the glowing, formal tattoos of his committees and lifebrands adorning his massive triceps. You swallow on a dry throat. Derya, of all people, hearing you scream!

He gives you a hooded look. "They infect themselves with some designer virus — it lets them hack the city's person-recognition systems. So the mites don't see them when they jump. Watch..."

She's swept past the whalelike oval of the public pool on the 202nd, past the sloping mandala of the Google offices on the 164th. At the 131st, just below her, is the old Stock Exchange, hung upside-down now as a hipster den.

Now the mites are finally closing in. A silver swarm coalesces around the 152nd,

and she vanishes into it like a snip of scallion into cloudy miso soup. When the cloud disperses, she's standing on one of the Stock Exchange's overhangs. She waves, antlike, then crawls through a dormer window.

"It's not funny," Derya says. "They're a huge drain on emergency preparedness. Ripple effects are causing project slowdowns..."

"Freeloaders drive systemic evolution," you find yourself saying.

"Don't you quote the founders at me," Derya snaps. "The Free Society is fragile. The minute enough people find anticontributive behaviour cool, the party's over — it's back to capitalist competition or state control." He stares until you meet his eye. "You even talk to those people — are you paying attention? — you

even talk to them, your rep will be trashed on all the major servers. You won't work, you won't party, you'll be defriended by every one of your tribes. Got it?"

Her broad-brimmed hat is still sailing on the wind. The mites missed it. It cuts between the towers of the 50th.

The upside-down trading floor is deserted. There are heaps of yellowed euros and deutschmarks dumped here, like snowdrifts. Wood panel, marble. Silence. And the air is strangely clear. You realize: no mites. The city has no eyes or ears, here. You walk through empty, miteless rooms, stepping around light fixtures.

Then she's there, in a doorway. Her eyes, bright blue, radiant. Her smile, with that chaste yellow dress, so bashful. She comes to you.

"You want it?" she says. "You want to be infected? You want to fly?"

You nod.

Eyes closing, she leans in for the kiss. ■ Benjamin Rosenbaum's fiction has appeared in Harper's, McSweeney's, Asimov's Science Fiction, Fantasy and Science Fiction and Strange Horizons. He lives among the tangled superhighways and horizontal strip malls of northern Virginia. His website is www.benjaminrosenbaum.com.

JACEY